OM nucleic - nucleic search, using sw model

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: May 3, 2003, 21:11:56 ; Search time 407.991 Seconds
                                              (without alignments)
                                              16426.693 Million cell updates/sec

Title: US-10-027-000-1
Perfect score: 2976
Sequence: 1 ttatagtcgcttgttaaat..........aaaaaaaaaaaaaaaaaaa 2976

Scoring table: IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0
Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   N_Geneseq_101002:*
  1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
  2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
  3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
  4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
  5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
  6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
  7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
  8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
  9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
 10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
 11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
 12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
 13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
 14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
 15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
 16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
 17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
 18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
 19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
 20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
 21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
 22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
 23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
 24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
      /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	260.8	8.8	588	21	AAF15000	Trichoderma reesei
2	184.4	6.2	3241	24	AAS18442	Contig 93 DNA enco
3	179	6.0	2401	21	AAA75634	Nucleotide sequenc
4	179	6.0	2401	21	AAZ56002	Contig 001 from co
5	177.4	6.0	1145	16	AAT04785	Cellobiase gene fr
6	175.2	6.0	2430	24	AAZ87294	S. venezuelae macr
7	175.2	5.9	2430	24	AAD39052	Streptomyces venez
8	175.2	5.9	12441	21	AAZ87284	S. venezuelae deso
9	175.2	5.9	13613	24	AAD39043	Streptomyces venez

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

ALIGNMENTS

10	167.2	5.6	13613	21	AAA87319
11	161	5.4	2256	17	AAT32999
12	138.6	4.7	50937	21	AAA09469
13	135.8	4.6	2166	18	AAT93682
14	135.8	4.6	2166	19	AAV36911
15	111.6	3.8	4403765	22	AAA199683
16	111.6	3.8	4411529	22	AAA199682
17	103	3.5	985	24	ABQ44816
18	103	3.5	985	24	ABQ44817
19	101.4	3.4	16836	19	AAV52831
20	98.6	3.3	3849	22	AAF25795
21	93.4	3.1	2271	22	AAN80309
22	91.2	3.1	2291	9	AAQ06844
23	85.8	2.9	12588	15	AAQ63293
24	85.8	2.9	65140	22	AAD17184
25	85.4	2.9	125401	22	AAD17186
26	84.4	2.8	390	13	AAQ21833
27	84.4	2.8	390	14	AAQ36859
28	84.4	2.8	390	22	AAF76910
29	84.4	2.8	390	24	AAI72775
30	83.8	2.8	3032	13	AAQ25325
31	83.8	2.8	3033	21	AAA63953
32	83.8	2.8	3033	21	AAZ06824
33	80.4	2.7	1140	19	AAV41733
34	80.4	2.7	7498	24	AAZ38879
35	80.2	2.7	1661	22	AAH21757
36	78.6	2.6	2064	14	AAQ52638
37	78.6	2.6	3269	16	AAT06024
38	78.2	2.6	1294	20	AAZ06824
39	78.2	2.6	1294	22	AAH74537
40	78.2	2.6	1294	16	AAQ85043
41	78.2	2.6	1521	17	AAT62676
42	78.2	2.6	1521	19	AAV11824
43	78.2	2.6	1647	16	AAQ85042
44	78.2	2.6	1647	17	AAT62675
45	78.2	2.6	1647	19	AAV11823

RESULT 1
AAF15000
AAF15000 AAF15000 standard; cDNA; 588 BP.
XX
ID
AC   AAF15000;
XX
DT   13-MAR-2001 (first entry)
XX
DE   Trichoderma reesei EST SEQ ID NO:7523.
XX
KW   Multiple gene expression; filamentous fungal cell; EST;
KW   expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW   Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW   culture condition; environmental stress; spore morphogenesis;
KW   metabolic pathway engineering; catabolic pathway engineering; ss.
OS   Trichoderma reesei.
XX
PN   WO200056762-A2.
XX
PD   28-SEP-2000.
XX
PF   22-MAR-2000; 2000WO-US07781.
XX
PR   22-MAR-1999;   99US-0273623.
XX
PA   (NOVO) NOVO NORDISK BIOTECH INC.
PA   (NOVO) NOVO NORDISK AS.
XX
PI   Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
DR   WPI; 2000-594572/56.

S. venezuelae deso
Chimaeric thermost
Streptococcus olea
Thermotoga maritim
Thermotoga maritim
Mycobacterium tube
Mycobacterium tube
Oligonucleotide fo
Oligonucleotide fo
S. chrysomallus ac
Acetobacter xylinu
Streptomyces nours
Streptomyces nours
Entire amylase gen
Sequence encoding
Amylase gene from
Sequence containin
Randomising oligon
PCR primer for 5'
Oligo #7 for cloni
Oligo #7 for cloni
Trichoderma reesei
DNA encoding a bet
Trichoderma reesei
Codon-optimised RA
Modified HIV prote
Corynebacterium gl
Streptomyces fradi
cDNA encoding aven
Streptomyces albid
Nucleotide sequenc
3-Hydroxysteroid-o
3-Hydroxysteroid o
Streptomyces sp. s
3-Hydroxysteroid-o
3-Hydroxysteroid o
Streptomyces sp. s

XX
PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
XX
PS Claim 89; Page 3038-3039; 3161pp; English.
XX
CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.
XX
SQ Sequence 588 BP; 130 A; 167 C; 157 G; 109 T; 25 other;

Query Match   8.8%;   Score 260.8;   DB 21;   Length 588;
Best Local Similarity   92.8%;   Pred. No. 3.3e-33;
Matches 283;   Conservative   0;   Mismatches   20;   Indels   2;   Gaps   1;

```
Qy  2208 ATCCAGGCCTGGTACGGCGGCAACAGACAGACGGGCAACTCCATTGCGACCTTCGTTCTTTGGC 2267
         |||||||||||||||| ||| ||| ||| |||||||||  || |||||| ||| || |||||||
Db     1 ATCCAGGCCTGGTACGCGGGNAACGAGACGGGCAACTACCATTGCCGACGTCGTCTTTGGC 60

Qy  2268 GACTACAACCCCTCGGGCAAGCTCGTGCCCTCAGCTTCCCCAAGCGCCCTGCAGGACANACCCC 2327
         |||||||||  ||| |||||||| |||||||||||||||||||||||||||||||||| |||||
Db    61 GACTACAACCCCTCGGGCAAGCTGTCCCCTCAGCTTCCCCAAGCGCCCTGCAGGACAACCC 120

Qy  2328 GCGGTTCTCAACTTCCGCACCGAGGCGCGGGCGCGGCCGGGACACGTCTACGTC 2387
         ||| |||||||||||||||||||| |||| || || ||||| | |||||||| ||
Db   121 GCGTTTCTCAACTTCCGCACCGAGCGCGGCGCACGCTGTACGGCGGAGGAGCGTCTACGGTC 180

Qy  2388 GGGTACAGGTACTACGGAGTTTGCCGACAAGGACGTCAAATTCCCCTTTGGCCACGGGCCTG 2447
         ||||||||||||||||||||||||||||||| |||| |||||||||||||||||||||| |||
Db   181 GGGTACAGGTACTACGGAGTTTGCCGACAAGAGCTCAATTTCCCCTTTGGCCACGGGCTG 240

Qy  2448 TCCTACACCACACTTTT--GCCTTTTCCAATCTCTCCGTGTCTCACANAGGCAAGCTCA 2505
         ||||||||||| ||||||  | ||||||||| ||||| ||| |||||| || || ||||
Db   241 TCCTACACCACTTTTTCCCTTTTCCAAATCTCTTCCGGGGGTTCTAACANNGGACGGNAAA 300

Qy  2506 GCGGTG 2510
         ||| ||
Db   301 GCGTG 305
```

RESULT 2
AAS18442
ID AAS18442 standard; DNA; 3241 BP.
XX
AC AAS18442;
XX
DT 12-MAR-2002 (first entry)
XX
DE Contig 93 DNA encoding S. narbonensis polyketide synthase.
XX

KW Narbonolide polyketide synthase; PKS; narbomycin modification enzyme;
KW erythromycin; rapamycin; tylosin; picromycin; methylmycin; neomethymycin;
KW agriculture; ds.
XX
OS Streptomyces narbonensis.
XX
PN US6303767-B1.
XX
PD 16-OCT-2001.
XX
PF 05-NOV-1999; 99US-0434288.
XX
PR 05-NOV-1999;
PR 05-NOV-1998; 98US-107093P.
PR 27-MAY-1999; 99US-0320878.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Betlach MC, McDaniel R;
XX
DR WPI; 2002-065495/09.
XX
PS Claim 1; Column 20-22; 24pp; English.
XX
CC The present invention relates to recombinant DNA vectors (cosmids).
CC that encode for the narbonolide polyketide synthase (PKS) enzyme and
CC various narbomycin modification enzymes from Streptomyces narbonensis.
CC The recombinant DNA vectors can be used to produce recombinant ketide
CC synthases and a variety of different polyketides (e.g. erythromycin,
CC rapamycin, tylosin, narbomycin, picromycin, methylmycin and
CC neomethymycin) for use in agriculture, medicine and health. The
CC recombinant vectors may be used to produce polyketides in relatively
CC high yields. AAS18432-AAS18443 represent contig DNA sequences that
CC encode for S. narbonensis PKS enzymes.
XX
SQ Sequence 3241 BP; 520 A; 1174 C; 1125 G; 422 T; 0 other;

Query Match   6.2%;   Score 184.4;   DB 24;   Length 3241;
Best Local Similarity   48.2%;   Pred. No. 6.8e-21;
Matches 1028;   Conservative   0;   Mismatches   956;   Indels   150;   Gaps   12;

```
Qy   544 CGCTCGGTCTCCACATTCAACCAAACTCTGCTCAAGAGGCAGGTAAGATGAATGGGCAAAG 603
         ||| |||| | ||| |  ||| || ||| || |  ||||||||| |||| ||| || ||||
Db  1151 CCCTGGCCACCACCTTCGTACGAGTCCATGGCCCGACAGCTACGGCGGGTCATGGGCCG 1210

Qy   604 AGGCCATCGCTAAGAGTGCGCCATCGTGATCCTCGGCGCCCGACTATCAACATGCAACGGCTCCC 663
         || ||| | |  || || | ||| | || ||  ||||||| || || ||||| | |||| ||||
Db  1211 ACGGACGCCGCCGCTGGGCCCAGGACAGACGACATTGGTTCTGGGCCCCATGGGCGTGC 1270

Qy   664 CTCTCGGTGGACGTGGCCTTCGAGTCGAATGGTGAGGATCCAGGTCCAGGCACACTTTT 723
         | | |||||| || || || ||||| || |||||||| || |||||||| | |||||||
Db  1271 CACACGGCGGGCGCCGGAACTACGAGCACTTCAGCGAGGACGACCCCTGGTTCTCGCGGCACG 1330

Qy   724 CTGCGGCGCTCTCATCCGCGGCCATTCAGAGCCACTGAGGTGCCAGGGTCCAGGCCTACCTTGG 783
         | | |||| ||| ||| || ||||| |||||| | | ||| ||| |||  ||||| || ||||
Db  1331 CGGTCGCCCAGATCAAAGGGCCATCCAGGGTGCCGGGGTCTGATGACCAGCGGCAGACACTTCG 1390

Qy   784 TGTGCCAATCGATCAGGAGGAGCAGCGGCCATCGAGTCGGTGCAGACANTCGTCACGGGAGCG 843
         ||| |||| | |||||||| || || || |||| || ||||||||||| |  |||||| |||
Db  1391 CGGGCCAACAACCAGGAGAACAACCGCGTTCAGCGTTCAGCGGACGACGC 1450

Qy   844 TCCGTGAAATCTACGCACTCCCGTTCCAGATTGCGTGCGGAGGACTCCAGCCGCGGGTCGT 903
         |||| ||| ||| || | ||||| ||||| || || | |||| |||||| || ||| | |
Db  1451 TCCGCGAGATCGAGTTCCCGGCGTTCGAG---GCGTCCTCGAAGGGCCGGCGCGGGCCTCCT 1507

Qy   904 TCATGACCGGCGTACAAATGGCCATCAAATGGCCTGTTCGTGCAGCGGAGAGAAACCCTAAATATCTTG 963
         ||| | ||| | || ||| ||| ||||| ||| | ||||| ||||| |||||| |||| |||||| | |
Db  1508 TCANGTGTGCCTATAACGGCGTCAAACGGCAAGCGCGTCGCGGCCAACGACGACGAGCTGTCTCA 1567

Qy   964 ATGGGATGCTTCGAAAGGAAATGGGGTTGGGATGGCCTAATCANGAGGACTGGTACGGCA 1023
```

```
Db  1568 ACAACGTGCTGCGCACCAGTGGGCTTCCAGGGTTCGTTGCTGGCTCGCCA 1627
              ||| ||  | | |  ||||| | | |  |  |||| |      |||| |||
Qy  1024 CATACAGTACCACACAGAAGCCCGTTGTGTGGCAGCGCCCGAGATGCCGGACTTCCAC 1083

Db  1628 C---CCCGGGCACGGACGGCGTCATCACCACCAGGAGAT----GGGACGTCGAG 1680
              |  |||||| |  |  |  |||  |  |  | |    |    ||| |   |
Qy  1084 GCTTCCGAGGAGAAACACTCAAGTTCAACGGAGAAAGCCCTTTATCCACGTCA 1143

Db  1681 CTCCCACCCGGCGACATCCCGCCGGGGGGCGAGCCCTTCTGGGTGACGCG 1740
              |||  | |||||| |  ||  |||| ||||  ||  |  |||| ||||
Qy  1144 TTGACCAGAGGGCTAGGGAAGTTCTTCAGTTGCTGTGCTGCCTCGGAGTGA 1203

Db  1741 CTGA-------------AGCAGGCCGTAGGG 1780
              |||             | |  |||| |||
Qy  1204 CGGAGAACGGCCCCGAGACAGTGTCAACACACCCCCGAAACGGCGCTCTCCTTCCGGA 1263

Db  1781 CCGTGACCGCGGTGGGCGGAGCCGCATCGTCAACCAGATGACAAGTTCGGTGCTCCTCG 1840
              |||||   |||| |||  |  |||  || |  |  | |  | | || |  |||  |
Qy  1264 AGGTTGGCAACGAGGGCATCGTGCTGCTGAAGAACGAGAAACAGTTCTGCCCTTGAGCA 1323

Db  1841 CGACTCCGGGCGCCCCGGCGTGACAAGGCGGGCTGTGGGGTGTCCGGCAAGG 1900
              | | ||| ||||  || |  |||  ||| ||| |||    ||| |  ||||
Qy  1324 AGAAGAAGAAGACGGCTGATTGTGCGGCCCCAACAGCAGCCACATACCACGGCCGGGAG 1383

Db  1901 TCGGCCGAGAACGGGCGGGGTGTCCTGCGCCAGCCACACGACTCGCCGGTG 1960
              || ||||| |  |||  ||| |||  | |  || |  |||| | |  ||||
Qy  1384 GCTCTCTGCGCACTCAGGGCCGCTACTAGCGTCTCCCTTTGACGGCGTCTCAGCACAGC 1443

Db  1961 ACGCCGGCGAAGAGCATCCGTGTCATCGGCGC-CCCAAGGTCACCGGC 2019
              |   |||||  |  ||  |  ||| |  |||| |  || ||| |  ||
Qy  1444 TCGAGACGACCGCCCCGCCATCGTTACACCGTCCGCGCTTCCTCCCATTCTAGGCG 1503

Db  2020 CTGGGCCACGCGCCCACGTCGTCGTCCCGGACTCGGGGCTCAGGGAGCCC 2079
              ||||||| ||   |  |  ||||||  ||| |||| |||  |  | ||||
Qy  1504 AGCAGTGCCTCACCGCCCCGACGGGCTCCGGTGCTCGGGGTGCTTCAACGAGCC 1563

Db  2080 CGGCGCGGGGCGGGGTCGTGAACGTACGAGACGGACGACCTTCGGGACGCGG 2139
              |||||  ||| ||| |||||| ||  |||| |||  || | ||||| |  |
Qy  1564 CTGGTACCCTACCGCCAGCACATTGACGAGCTCTTCTTCACCAAGAGCGACC 1623

Db  2140 ATCCCGGCCGGGGCCGCGCAGCCTCAGC-----CCGGCGTTCAACC 2173
              | |||||| |||| |  ||||||  |||     ||||| ||| |||
Qy  1624 TGGTGGGACTACTACCACCCCAAGGCTGGCTGGACGCTGTACCGCACGT 1683

Db  2174 AGGGCCACCAGCTGGAGCGCGGGGCAAAGCCGCCCTGTCCTGCCGCGACGT 2233
              ||||||   ||||||||  |||||  ||| ||| |  |  ||| ||| |||
Qy  1684 ACACCGCCGACGAGGGACTGCCACCCTACAGACTCGGTGCTCGTGCGGCGAAAGG 1743

Db  2234 CCCGCCGACGGGCGAGTACCGCACATTCGGTGCTGACCGGCGTGCAG- 2292
              |||||| |||||  ||||| || ||   ||| || |  ||| ||| |
Qy  1744 CGTACGTAGACGACCAGCTCGTTCGTGTCGACAACGCCACCAAAGACAGGGTCGCCT 1803

Db  2293 -------------CTCGGCCAGCCACA 2305
                           || |||||||||||
Qy  1804 TCTTCGGCTCCGCGCCACCCGCGGAGAGAGGCGGCGCCATCAAATCTCGTTCAAGGGCAACACGT 1863

Db  2306 CCATCGAGCCGCCGGTCAGGTCTACGGCAAGGTGACCAGCGCCAAAG 2365
              || | ||||| || |  |||||| |||||||| | |  |||||| ||
Qy  1864 ACAAGTTCAAGATCGAGTTCGGGCTCCGCACCACTACACCCTCAAGGGCGACACCATCG 1923

Db  2366 GCACGCACAAGCT------CACGACTCGGCGTTCGCGATGA 2401
              |||  | |||| |      | || ||| ||  | ||| ||
Qy  1924 TCCCGGCCACGGCGTCCTCTCCGGGTCGGCGGGTTGCGGGCGACACCATCG 1983

Db  2402 GCGGCCACCGCGCTCCCCTGGAGCTGGTGGGCTGCAAAGGGCAACCGCAGA 2461
              ||||||||||  | | |  |||| ||| | |  | ||| || || | ||
Qy  1984 TCGAAAAGTCCGTCGCCCTTCGCCAAAGGACCCAGGGTTGAAGGTCATTGACGAACCGGAA 2043

Db  2462 TCGGGAAGGCGCTGGGAGTCGGCGGAAGGCCCGTACGGCAAT------2514
              |  ||||||| |||| | | || |   |||||  | | | |
Qy  2044 ACGCGGACTGGGAGGACGAGGGGCCGAGCGCGGGGGCGATGAGCTCCCCGGCCGGTGCTGG 2103

Db  2515 --TACGACGACGCACCGAGGGCGTGGACCGTCCGGACCGTCCGGACCGTCGTCTCGGTTGCTGCCGGGTACGGCAGG 2572
              | |||| | ||  |  |  || |   || |      |||||| |||  | || || | | ||||  |||
Qy  2104 ACCAGCTCATTGCCGACGTGGCCGCGGCGCGAGGTGGCGGACGCCAAAACACCACGGTCGTCGCCTGCGGGTACGCGG 2163

Db  2573 ACAAGCTGGATCTCGGCGCCGTCGGCCAGCCTCGGAGATGCCGGACCAGATCGGTGTTCAGCCGTCATCCAGGCTGGTACG 2632
              |  ||||| | || | | |||| | || ||  |  |  | | |  | || || ||| || || || |   || ||  |||
Qy  2164 GCACCCCGGAGGAGATGTCGCTGGTCGAGGTGGGCGGTGTCGGCTCGACGCCCACGGCGAGAACCCGGACGGTGGTACC 2223

Db  2633 GTTCGTGCTGGGTGCGTGCTGATCGCTGTCCAAGACCCCGCGGTCCTTGGACGTCCAGGACGG 2692
              || | | | || | || || | ||  |  ||| | ||||| ||  || ||| || || | ||
Qy  2224 GCGGCAACGACGAGCGGGCAACTCCATTGCCGACGTGCGTCCTTTGTGGGCGTCCTTGGACGTCACCCCTCGG 2283

Db  2693 CGGGCGACTCACCGGAGGCGACCGGGCCCAGGCCGGCGGAGGCGTCGACGGCGGAACCCCGCGG 2752
              | || | || ||| |||| ||| || | ||| ||| | || || | || || |  || || ||
Qy  2284 GCAAGCTGTCCCTCAGCTTCC-----CCAAGCGCCGGGGCGCTGCTCTACGGTGAGACCCGTGCG 2337

Db  2753 GCAAGCTCACCGACAGACGCTTCCGGAGCCGCGACCGGCGTTCTGCCTGGTGACGTGAACCCGAAGG 2812
              ||||||| ||||| | | |  ||  || || || || | | |  |  |||| |||||| |  ||||
Qy  2338 ACGCGTACCCGGAGGCGCACGCTTT--GTACGGCGAGGCCAGCAGCTCCACGCTGACGGTCTACGGGGTACA 2394

Db  2813 ACCGCTACCCGGCGGTCGACAACAGCAGAACGTACACCAGGCATCCAACGTGGTACGTGGGGTACC 2872
              || || ||| | |  | || |  |  ||| || ||||| | |   || ||| |||| | ||| ||
Qy  2395 GGTACTACGAGTTTGCCGACAAGGAGAAACCGGCGTCAATTTCCCCTTTGCCGACAGGCGCCCGTGTCTCCTGTCCTTGGGGTACA 2454

Db  2873 GCTGGTTCGACAAGGAGAAACGGCCGTCGAGCGCGGTCGCAAGGAGTCACCGTTCGGGGCACCTCTACGTCA 2932
              | |||| |||||| || ||| ||  | || || || || || |   | || ||  || ||| |  ||| ||
Qy  2455 CCACTTTGCCTTTTCCAAATCTTCTCCGTGTCTTCACAAGGAC----GGCAAGTGAGGTGT 2511

Db  2933 CGTCGTTCACGCAGAGCGCCGCAGAGCGCCCCGACCCCGTTCCGGTTCGCCGACAGGCCGGGCACCCCTGTCGCCCGTTCGCCGTCA 2992
              ||| ||  ||||| || |||  |||  |  || |  |||| ||| | ||| ||| ||| || ||   ||||  || ||  ||| ||
Qy  2512 CCCTCTCCGTGAAGAACACCTGCTGGTCGTTCGCGCCTGTCGTTGGCGCCGCGTTTCCGGTTGGGCCGAGCTCGGTGAAGAAG 2571

Db  2993 CGGTCACGGTGCGCAACAGCGGGGAACGGCGGCCCGACGTGGTGTCCAGGTGCCGTTCGCCAGGTCTACGTCG 3052
              | || ||| | || || | | || ||||  | | || | |||| |  || | ||   | |||| ||| | ||
Qy  2572 AGCCCCTTCCAAAGCGCCAAAGATTAAACCCCCCGTCAAGGGGGTCAAGGGGTCGTCCAGGGGGTATCTCG 2631

Db  3053 GCCGCGAGCCCGAGAAGGTGACGGCGCTCCGGGAGAAGCTGTCGGTGGGCTTACACGAAGG 3112
              | |||| || | || ||||  || ||  ||| |||| || || || |||  |||| ||||
Qy  2632 TCGAACTGCAGCCCCGGCGGCGGGGCCGCCGAGGGTGACGGCGGGCTCAGTCGAAGACGGTGAC 2665

Db  3113 TCGGCGCTGCGCGGGGCCGGGCAGTCGAAGAGTCGAAAGCGGTGAC 3146
```

RESULT 3
AAA75634

```
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Ashley G, Betlach MC, Betlach M, Tang L, McDaniel R;
XX
DR   WPI; 2000-610844/58.
XX
PT   New recombinant pick hydroxylase gene of Streptomyces venezuelae useful
PT   for converting ketolides to antibiotics and as antibiotics and
PT   intermediates in the synthesis of compounds with pharmaceutical value
PS   Disclosure; Columns 39-40; 117pp; English.
XX
CC   The present sequence is used to produce the recombinant DNA compounds
CC   of the invention. The specification describes a recombinant DNA compound
CC   expressing recombinant polyketide synthase genes in host cells for the
CC   production of narbonolide, narbonolide derivatives and polyketides that
CC   are useful as antibiotics and as intermediates in the synthesis of
CC   compounds with pharmaceutical value. The DNA compounds may also encode
CC   a C12-hydroxylase (pickC), desosamine biosynthesis and, desosaminyl
CC   transferase enzymes (useful for conversion of ketolides to antibiotics),
CC   and the beta-glucosidase enzyme (involved in picromycin biosynthesis).
CC   These compounds are also useful for increasing the antibiotic activity
CC   of a compound relative to the unhydroxylated compound. The recombinant
CC   host cells are useful as genetic systems that allow rapid engineering
CC   of the narbonolide polyketide synthase. These would be valuable for
CC   creating novel ketolide analogs for pharmaceutical applications.
XX
SQ   Sequence 2401 BP; 405 A; 859 C; 828 G; 308 T; 1 other;
```

Query Match   6.0%; Score 179; DB 21; Length 2401;
Best Local Similarity   47.7%;   Pred. No. 4.9e-20;
Matches 1018; Conservative   0; Mismatches 966; Indels 150; Gaps 11;

```
Qy   544  CGCTCGGTTCCACATTCAACCACAAACTCTCCTTCGAAGAGGCAGGTTAAGAGATGAGTCGGGCAAAG  603
          || |   |  ||   ||||   ||| |    |  ||||     |||   ||    | || ||||
Db   255  CCCTGCCCAGCACTTCGACGACACCATGCCCAGCGTACGGCAAGGTCATGGCGCGG             314

Qy   604  AGGCCATCGCTAAGAGTGCCCATGTGATCCTCGGCCGGCGACTATCAACACATGCAACGCCTCCC    663
          |||   ||      ||| || |  ||  |||||||  |||  |  |||||| ||    |  ||||
Db   315  ACGGTCGCGCGGCTCAACCAGGACACATGGTCCTCGGCCCGATTGATGAAACAACATCCGGGTGC    374

Qy   664  CTCTCCGTCGGACGTGGCTTTGCAGTCGGTGCGTTCCTGGGCGGTTGGGAG                 723
          || |  || ||  | || |  ||||   ||| | |||  ||||||||| ||
Db   375  CGCACGGGCCCGGAACTACGAGACCTTCAGCGAGGACCCCCGGGGCTTGGGAG               434

Qy   724  CTGCGGCTCTCATCCGCGGCATTCAGAGAGCACTTCAGCGTACGAGTCAACGACTTT           783
          ||| |  |  | ||| ||||||||||||||| |||||  | |||||||||||| ||
Db   435  CGGTCGCCCAGAATCAAGGGCATCCAGGGTCGGTCTGATGACCACGGGCCACGTCG           494

Qy   784  TGTGCAATGATCAGGAGCGACAGGGGCCATGATAGTTGCAGAGCATCGTTCACGGAGCGGGGAG     843
          |||   ||| ||   |||| |  ||||||  |  || |||||| |  |   |  || ||||||
Db   495  CGGCCAACAACCAGGAGAACACCGCTTCTCCGTGAACGCCAATGTCGACGACGATGA          554

Qy   844  TCCGTGAAATCTACGCACTCCCGTTCCAGATTGCTTGTGCGAGAGTCCCAGCCGCGGTGCCT      903
          | || |   | |  ||  || || ||   |||  || | |||||||| |||||| ||||||
Db   555  TCCGGCGAGATCGAGTTCCCGGCGTTCGAG---GGCGTCCTCCAAGGGCCGGCGGCGTCCT     611

Qy   904  TCATGACGCGCTACACAGCGACATCAAATGGCCATGAATGGCGTGTTCGTGCGAGCGCCAGAACC   963
          ||    ||| |||   ||   |||     | |||  | || ||| | |||| || || || |  |
Db   612  TCATGTGTGCCTACAACGGGCCGCTCAACGGGAAAGCCGTGCTGAGGTCGCCGCAAGCGCTCTCA   671

Qy   964  ATGGGATGGCTTCGAAAGGAATGGGGTTGGGATGGCCTAATCATGAGCGACTGGTTACGGCA      1023
          ||| || |||  ||  |||  |||||| | ||   || || ||  | |  | ||    |||
Db   672  ACAACGTGCTGCCCACGCACGAGTGGGGCTTCCAGCGCTGGGTCGACGTCGCGGTGACCGCAGG    731

Qy   1024 CATACAGTACCACAGAAGCGCGTTGTTGGCAGGCGTTCGACCTGGGGACCTCCAC             1083
          ||     | | ||  ||  | | |||   ||  |||||| ||||||||||| |||
Db   732  C----CCGGGCCACGGACACCGGAGGTCCTCAACCCGAACACG                        775

Qy   1084 GCTTCCGAGGAGAAAACTCAAGTTCAACGTCTTCCCAACGGAAGCCCTTTATCCACGTCA       1143
          ||     | |||   || |   |   |  | || || ||||  || |  |  | ||||||
```

```
Db   776  ------GGCGTCGAGCTCCCCGCGGCGACCTCCCCGGAAGGGCGAGCCCTGGCCGGGCCA     827
                |||  || |||||||||||||||||||||||||||| ||||||| ||||||||||
Qy   828  AGTTCTTCGGCGA---GGCGCTGGAAGACGGCGCCGTCCTGAAGACAGTGTGCTGCTTGCCGAGGTG  884

Db   945  CCACTCCGGCGCCCGCGGCCCGAGCCGCGACAAGGCGGTGCCCGCAAGG                  1004
          ||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1264 AGGTTGGCAACGAAGGGCCATCGTGCTGAAGACGAACGAGGAGAACAAGGTTCTGCCCGTTGAGCA  1323

Db   1005 TCGCCGGAAACGGCGCGGGTGCTCCTGCGGCGAACGGGGGTCTGCGCGGGTG             1064
          |||||||||||||||||||||||||||||| || |||||| ||||| ||||
Qy   1324 AGAAGAAGAAGAACGGCTGAGTTGTCGGCCCCCAACGCCAGGCCCACATACCACGGCGGAG       1383

Db   1065 ACGGCGGCCAAGAGACAATCGCGGGTATACGGCCGGCGTCATCGGGCCGGC-CCCAAGGTCACCGGC  1123
          ||||||||| |||| | ||||  | ||||||  ||||||||| ||||| || ||||| ||||||||
Qy   1384 GCTCTGCCCGCACTCAGGGCCTACTACGACGAGTCACTCCCGTTTGACGCGCGTTCCAGCGAAAGGC   1443

Db   1124 CTGGGCAGCCGGCCCCAGTCGTTCCCGGACTCGGCGGCCGCCCACTGCGACAACCATCAAGGCC     1183
          ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Qy   1444 TCGAGACGCGCCCATCGTACACGTCGGGCCTACCACCACCGGTCCTCCTCCCATTTCCTAGGCG     1503

Db   1244 ATCCCGCGGCGGGGGGAACCTCAGC------CCGCGGTTCAAC                        1277
          ||||||||||||||||||||||||        |||||| ||||
Qy   1564 CTGGTAGCCCCTAACCGCCAGCACACATTGACGAGAGCTTCTTCACCAAAGACGACAATGCCC    1623

Db   1184 CGCCGCGGGGGTGCGGGGTGACGGTGACGTACGGACGGGGTGAGGAGAGACCTTCGAGGACCAG   1243
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1184 CGCCGCGGGGGTGCGGGGTGACGGTGACGTACGGACGGGGTGAGGAGAGACCTTCGAGGACCAG   1243
```

```
Qy  2224  GCCGCAACGAGAGGGGCGAACTCCATTGCCGACGTCGTCTTTGGCGACTACAACCCCTCGG  2283
          |||| ||| |    |||                           |     |||  |||||
Db  1797  CGGGCGCAGGCGGCGGGCGCCGAGGCCACGGCCTGCTCTTACGGTGACGTCAACCCGAGCG  1856

Qy  2284  GCAAGCGTGTCCCTCAGCTTCCC------------AAGGCCGTGCAGGACAACCCCGGGTTTC  2334
          |||||| |  | |||  | |||||            |      |||  |  |     ||  |
Db  1857  GCAAGCTCACGCAGAGACGTTCCCGCGCCGCCGGAGAACCACGCGGTCGCGCGGGACCCGA  1916

Qy  2335  TCAACTTCCGCACCGAGGCCCGGGACGCTGTACGGCGGAGGACGTCTACGTCGGGTACA  2394
          | | |      |  |  |   |     | |   ||  || |  |||  |  |||||||||
Db  1917  CCAGCTACCCGGGCGTCGACAACCAGCAGACGAGACGTACCGCGAGGGCATCCAGTCGGGTACC  1976

Qy  2395  GGTACTACGAGTTTGCCGACAAGGAGAGTCAATTTCCCTTTGGGCACGGCCTGTTCCTACA  2454
          | | |  |||   |||||||||||  |   | ||    |  || |||||||||   | |||
Db  1977  GCTGGTTCGACAAGGAGAACGTCAAGGCGCGTGTTCCGCTTCCGGCCACGGCCTTCGTTACA  2036

Qy  2455  CCACTTTTGCCTTTTCCAAATCTCTCCGGTGTCTCACAAGGAC---GGCAAGCTGAGCGGTGT  2511
          || |   | ||      ||  | |  |  |||  |  ||| || |  |||  |||  |||||||
Db  2037  CCTCGGTTCACGCAGAGCGCCCCGACCGTGCGGTGCTGCTCACGGTGGTTCTGAAGGTCA  2096

Qy  2512  CCCTCTCCGTGAAGAACACCGGCTGCTCCGTGTGCCCGGGCGACAGTGGGCCCAGCTCTACGTCA  2571
          || | |||||      |||  ||| ||    ||  ||||||| || ||||| | |||  ||||||
Db  2097  CGGTCAGGGTCCGCCCAACAGCGGGAAGCGCGCGCCCGGGCCACGAGTGGCCGTCCAGGCGTACCTCG  2156

Qy  2572  AGCCCCTCCAAGCGGCCCAAGAGATTAAACCGCGCCCGTCAAGGAGGTCAAGGCCTTCGCAAAGG  2631
          ||||  |||||| |  || || |    |  |||||  ||||| || ||| |   || |  |||||
Db  2157  GTGCGCAGCCCGAACGTGACCGGCTCCGCAGGCGAAGAAGAAGAGCTTCGTCGGGGCTACACGAAGG  2216

Qy  2632  TCGAACTGCAGCCCCGGCGCGCGAGCGAGCGAACGGGCGGTGAC  2665
          || | ||| || || || |  | |||| |||  || ||||||||
Db  2217  TCTCCGTCCGCGCCGCGCGCGGAGCGGCGGCGAAGACGGTGAC  2250
```

RESULT 4
AA256002

ID   AA256002 standard; DNA; 2401 BP.
XX
AC   AA256002;
XX
DT   23-MAR-2000 (first entry)
XX
DE   Contig 001 from cosmid pKOS023-27 from Streptomyces venezuelae.
XX
KW   Narbonolide polyketide synthase; PKS; cosmid pKOS023-27; contig 001;
KW   ketolide; 1 beta glucosidase; antibiotic production; narbomycin;
KW   picromycin; ds.
XX
OS   Streptomyces venezuelae.
XX
FH   Key   Location/Qualifiers
FT   CDS   80..2389
FT   /*tag= a
FT   /product= 1_beta_glucosidase
XX
PN   WO9961599-A2.
XX
PD   02-DEC-1999.
XX
PF   27-MAY-1999;
XX
PR   28-MAY-1998;   98US-0087080.
PR   28-AUG-1998;   98US-0141908.
PR   22-SEP-1998;   98US-0100880.
PR   08-FEB-1999;   99US-0119139.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Ashley G, Betlach MC, Betlach M, McDaniel R, Tang L;
XX
DR   WPI; 2000-072618/06.
DR   P-PSDB; AAY67212.
XX

PT   New recombinant DNA encoding a domain of narbonolide polyketide
PT   synthase, for production of ketolide antibiotics -
XX
PS   Disclosure; Page 34-35; 98pp; English.
XX

CC   This is contig 001 from the recombinant cosmid pKOS023-27 DNA sequence
CC   (see AA256001) which contains a Streptomyces venezuelae DNA insert. The
CC   cosmid contains open reading frames which encode the various modules of
CC   the narbonolide polyketide synthase (PKS). The invention relates to
CC   recombinant DNA containing a coding sequence for a narbonolide PKS.
CC   Polyketides are compounds synthesised from 2-carbon units through a
CC   series of condensations and subsequent modifications. Modular PKSs are
CC   responsible for the production of many antibiotics including picromycin.
CC   The narbonolide PKS consists of a loading module, six extender modules,
CC   and two thioester domains. Four proteins make up the narbonolide PKS
CC   (PICAI, PICAII, PICAIII and PICAIV). PICAI includes the loading module
CC   and extender modules 1 and 2, PICAII includes extender modules 3 and 4,
CC   PICAIII includes extender module 5 and PICAIV includes extender module 6
CC   and a type II thioester domain. The second type II thioesterase
CC   domain is found on the PICB protein. The nucleotide sequences encoding
CC   all of these proteins can be isolated in recombinant form from the
CC   recombinant cosmid pKOS023-27. Narbonolide is desosaminylated in S.
CC   venezuelae to yield narbomycin, the desosaminyl transferase enzyme is
CC   required for this conversion, and the desosamine biosynthetic genes are
CC   also found in cosmid pKOS023-27. The recombinant DNA of the invention is
CC   used to express, in transformed cells, narbonolide (or its derivatives)
CC   or other ketolides (particularly hybrids), which may then be converted
CC   (e.g. by other enzymes recombinantly expressed in the same hosts) to
CC   polyketide antibiotics or their intermediates. The antibiotics are useful
CC   in human or veterinary medicine.
XX
SQ   Sequence 2401 BP; 405 A; 859 C; 828 G; 308 T; 1 other;

Query Match   6.0%; Score 179; DB 21; Length 2401;
Best Local Similarity   47.7%;
Matches 1018; Conservative   0; Mismatches 966; Indels 150; Gaps 11;

```
Qy  544  CGCTCGGTTCCACATTCAACCAAACTCTGCTCCGAAGAGGCAGGTAAGATGATGGCCAAG  603
         | || |   ||  ||||   |     |||     ||||| ||||||  | | |||| |||
Db  255  CCCTGGCCCAGCACCTTCGACGACGCCGACAGCTCACGGCCAAGGTCATGGTCGCCGCG  314

Qy  604  AGGCCCATCGCTAAGAGTGCCGCATGTGATCCTCGGCCCGACTATCAACATGCAACGCTCCC  663
         | ||  |  | ||     |||||       | |||||||||| |||| |||| |||   |||
Db  315  ACGGTCGCGGCGTCAACCAGGACATGGTCTCGGCGCCCGATGAGTGAACAACATCGGGGTGC  374

Qy  664  CTCTCGGTGGACGTGGCGTTCGAGTCGAATTGGTTGAGAGATCCGTTCCTGGCGCGGCTTGGGAG  723
         | || |  ||  | |||  | ||  |||    |  |||||  |||||| || ||||    ||||
Db  375  CGCACGCGCGGCCGGAACTACGAGACACCTTCAGCGAGGAGAACCCCCTGGGTGTCTCCTTCGCGCACCG  434

Qy  724  CTGCGGCGTCTCATCCGCGGGCATTCAGGACGACTGGAGTGCAGGCTACGATCAAAGCACTTTT  783
         | ||| || || ||| |||| ||||||| | | |||  |||||| |||||||| || ||| |
Db  435  CGGTCGCCCAGATCAAGGGCATCCAGGGTGCGGGTCTCGATGACCACGGCCAAGCACTTCG  494

Qy  784  TGTGCAAATGATCAGGAGGACCAGGCGCATGATGATGGTGCAGAGCATCGTCACGGAGCGGGGCTC  843
         ||||| |  |||||||| || ||||| ||| | ||||| ||||||||||| || ||||  ||| |
Db  495  CGGCCAACAACCAGGAGAACAACACCGCTTCTCCGTGAACGCCAATGTCGACCAGCAGACGC  554

Qy  844  TCCGTGAAATCTACGCACTCCCCGTTCCAGATTGCTGTGCGAGACTCCCAGCGCGGGTTGGCGT  903
         |||   || | | |  | |||||||||| | | | ||| |||  |||||  ||| || | ||
Db  555  TCCCGCAGATCGAGTTCCCGGCCGTTCGAG---GCGTCCTCCAAGGCCCGGGCCCGGCGGCCT  611

Qy  904  TCATGACGGCGTACAATGGCATCAAATGGCCGTGTCGTGCGAGCGAAGAACCCTAAATATCTTG  963
         ||| | |  ||||||| || || ||  ||||| | |||  ||| ||  | || | |  | |
Db  612  TCACGTGTGCCTACAACGGCCTCAACGGGAAGCCGTCCTGCGCAACGACGAGCAGCTCCTCA  671

Qy  964  ATGGGATGGCTTCGAAAGGAATAGGGGTTGGGATGGCCTAATCATGAGGCGGCACTTGTACGGCA  1023
         |  | |||||| || |||| || | || |||| ||||  || || || ||||  ||| ||||||
Db  672  ACAACGTGCTGCCGCACGACAGCAGTGGGCCTTCCAGGCGTGGGGTGTCCGGACTTGGCTCGCCA  731

Qy  1024 CATACAGTACCACAGAAAGCCGTTGTGTTGGCAGGCCCTCGACCTCAGCAAGGATGCCGGGTGCGT  1083
         || |    |||||| |  ||| |   | || ||||||||||||| || |||||| |
Db  732  C---CCCGGCCACCGACGCCCATCACCAAGGGCCTCGACCAGGAGATG------------  775
```

```
Qy   1084  GCTTCCGAGGAGAAACACTCAAGTTCAACGTCTCCAACGTCCCTTTATCCACGTCA 1143
Db    776  ------------GGCGTCGAGCTCCCCGGCGAGACGTCCCGGAGAAAAGCCGGAAAAGCCCTTCGCGCGCGGCCCA 827
           |||   ||||

Qy   1144  TTGACCAGAGGGCTAGGGAAGTTCTTCAGTTGGTCGTCAAGAAGTGTGCCTCCGGGAGTGA 1203
Db    828  AGTTCTTCGGGCGA---GGCGCTGAAGACGGGCGCCGTTGCCTTGAAGACGGCACGGTCGCCCCGAGGCGCG 884

Qy   1204  CGGAGAACGGCCCCGAGACGACGACTGTCAACACACCCGAAAACGGCCAGCTTCCTCCTTCCGGA 1263
Db    885  CCGTGACCGCGGTCGGGCCGGAGCCGGATCGTCGCCGATGGGACGAAGTTCGGGTCGCTGCGCTCCTCG 944

Qy   1264  AGGTTGGCAACGAGGGCCATCGTCGCTGAGAAGAACGACACAACGTTCTGCCCTTGAGCA 1323
Db    945  CCACTCCGGCCCCGAGCCGGGCGCCCCGAGCCGGCGTGCCCGCGCAAGG 1004

Qy   1324  AGAAGAAGAAGAGCTTGATTGTGCGGCCCCAACGGCCACACATACCACGCGGGAG 1383
Db   1005  TCGCCGAGAACGGCGGGCGGTGCTTCCTCGCGAAGCAGGGCCCTGCCGGTGCCGGGTG 1064

Qy   1384  GCTCTGCCGCACTCAGGGCCTTACTACGGACTCACTCCCCTTTGAGGGCCTTCAGCAAGCAGC 1443
Db   1065  ACGCCGGCAAGAGACAATCGCGGGTCATCGGGCCCGAC--CCCAAGGTCACCGCG 1123

Qy   1444  TCGAGACGGCCGCCATCGTACACCGTCGGCGCCTACACCACCGGTTCCTCCCATTCTTAGGCG 1503
Db   1124  CTGGGCAGCCGCCGCCACGTCGTCCGGGACTCGGCGCGGCGCCCACTCGACGCCATCAAGGCC 1183

Qy   1504  AGCAGTGCCTCACGCCGCGAGCGGCGGCTCCGGGCATGGTCGTTCAACGAGAGCCCC 1563
Db   1184  CGCGCGGGTGGGTGCGGACGGTGACGTACGAGACGGTTCGGAGGAGAACCTTCGGACGCCAG 1243

Qy   1564  CTGGTACCCCTTAACCGCACAGTGACGAGTCTCTTCTTCACCAAGACGGACACGCACC 1623
Db   1244  ATCCCGGCGGGGGAACCTCAGC-------CCGGCGTTCAACC 1277

Qy   1624  TGGTGGGACTACTACACCCCCAAGGGGGCAGACACGTGGTACGCCGACATGGAGGGCACGT 1683
Db   1278  AGGGCCACCAGCTCGAGCCGCGAAGGCGGGCGGCCGTGTACGACGCGTCGACCGTGC 1337

Qy   1684  ACACCGCCGACGAGGACTGCACTCTACGAGCTCGGCCTCGTCTGTCTCGCGGGCACGGGCAAAGG 1743
Db   1338  CCGCCGACGGCGGGCGAGTACCCGCCATCGCACCGGTGTTACCGCCACGGTGCAG- 1396

Qy   1744  CGTACGTAGACGAACCAGAGTCGTCGTCGACAACGCCCAAGAAGGTCCCCGGCGGATGCCT 1803
Db   1397  -------------CTCGCAGCCACA 1409

Qy   1804  TCTTCGGTCTCCGGCCACCCGCGAGGAGGCCGGCCATCAAATCTCGTCAAGGGCAACACGT 1863
Db   1410  CCATCGAGCGCGGTCAGGTCTACGGCAAGGTGAGCAC--CCCGTCCTCTCAAAGTGACCAAGG 1469

Qy   1864  ACAAGTTCAAGATCGAGTTCGGGCTTCGGGCTCGGCCGACCACTACACCCTACACCCTCAGAGGGCGACACCATCG 1923
Db   1470  GCACGCACAAGCT------------CACGATCTCGGCTTCGGGCTTCGCGGATGA 1505

Qy   1924  TCCCGGCCCACGGCGCTCCCTCCGGCGTCGCAAGGTCATTGACGACCAGCCGAAA 1983
Db   1506  GTGCCACCCGCGCTCTCCCTGGGACTCGGCGTNACGCGGGCGGACGACGA 1565

Qy   1984  TCGAAAAGTCCCTGCCCTCGCCCTACGGCAGCACGACAGCAGGTCATCATCTCGCGGGGGCCTTA 2043
Db   1566  TCCGGCGAAGGCCGTGGAGTGCGGCGGAAGGCCGTGGGCTACGACGATCGTCGCGGCGGTCAAGG 1617

Qy   2044  ACGCCGACTGGGGAGACCGAGGGCGGCGCGCCAGCATGAAGTCTCCCGGCGTGCTGG 2103
Db   1618  -CTACGACGAGGCGCACCGAGGCGGTCGACCGGTCGGGGCGTGCTCGG 1676

Qy   2104  ACCAGGTCATTGCCGACGTGGGCGCGCGAACCAACACCGTGCTGCCGTCGTCATGCGAGACGG 2163
Db   1677  ACAAGCTGATCTCGGCTGTGTCCGGAGACCCGAACCCAACACGACGAATCGTGGTGTCCTCAAACACG 1736

Qy   2164  GCACCCCGAGGAGATGCCCTGGCTCGAGCGCACGCCGCGTCATCCAGGACCTGGTACG 2223
Db   1737  GTTCGTCGGTGCTGCGTCGATCGCGAGCCTGGTCCTGTTCCAAGACACCGGCGGTCCGTCCTGGACTACGTTGGTACC 1796

Qy   2224  GCGGCAACGAGGACGGCGCAACTCCATTGGCCGACGGTCGTTTGGGCGACTACAAAACCCTCGG 2283
Db   1797  CGGGCGCAGGCGCGCCGAGGCGCCACCGGCTGCGTCTTCCTCTACGGTGAACGTCAAACCCGAGCG 1856

Qy   2284  GCAAGGCTGTCCCTCAGCTTCCCC--------AAGCCGCTCAGAGAACAACCCCGCGTTTC 2334
Db   1857  GCAAGCTCACGCAGAGCTTCCCGCGCGCGCGCGGAGAACCAGCACGAGCCACCGGGTCGCCGCGGCGACCGGA 1916

Qy   2335  TCAACTTCCGCACCGAGCGCGGCGGCCACGCGTCGTCAGGCGGAGGCTTCTACGTCGGGGTACA 2394
Db   1917  CCAGCTACCGGGCGTCGACAACCACGAGAGCGACGCGGTTCTCACGGCGGCATTCCACGTCGGGTACC 1976

Qy   2395  GGTRACTACGAGTTTGCCGACAAGGAGACGTCAAATTCCCCTTTGGGCCACGGACGGTGTGCCTCACA 2454
Db   1977  GCTGGTTCGACAAGGAGAACGTCAACCGCGTGTCCTTCCGTTCCGGGACAGGCGGCCTTACCGTACA 2036

Qy   2455  CCACTTTTGCCTTTTCCAAATCTCTCCGGTGTCTCAGCAAGGAC---GGCAAGCTGAGCGTGT 2511
Db   2037  CCTCGTTCACGCAGCAGATGCGCCCCGACCGGTCGTCGCCGTCACCGGGGTGGGTTCGTTGAGAGGTCA 2096

Qy   2512  CCCTCTCCGTGAAGAACACCGGCTCGTGCCCGGCCGCACAGGTCCCCAGCACTCTACGGTCA 2571
Db   2097  CGGTCACGGTCCGCAACAGCGGGAAAGCGGAGTCGTTCCAGCCGTTACCTCG 2156

Qy   2572  AGCCCCTCCAAGCGGCGCAAAGATTAACCGCCCGTCAAGGAGCTCAAAGGGGCTTCGCAAAGG 2631
Db   2157  GTGCCAGCCGAGAACGTGACGGCGCTCCGCAGCCGCGAACGAAGAAGCTCGTGGGCTACACGAAGG 2216

Qy   2632  TCGAACTGCAGCCCCGGCGCGACGAAGGCCGGTGAC 2665
Db   2217  TCTCGCTGCGCGCGGCGGCGGGCGAGGCGAAGACGGTGAC 2250
```

PS   Disclosure; Fig.14; 4lpp; English.
XX
CC   The 0.75-kb PstI and the 3.05-kb Pst-NdeI fragments of C. biazotea
CC   ATCC 486 chromosomal DNA were sequenced using the dideoxy method.
CC   Sections of the coding strand were identified, the first 5 of which
CC   were contiguous (given in AAt04785) and included a putative start
CC   codon. The sequences of the other 3, non-contiguous, downstream
CC   fragments are given in AAQ73991-93. Cloning of the gene region allowed
CC   extracellular prodn. of recombinant cellobiase in E. coli and S.
CC   cerevisiae.
XX
SQ   Sequence 1145 BP; 152 A; 444 C; 391 G; 158 T; 0 other;

Query Match   6.0%; Score 177.4; DB 16; Length 1145;
Best Local Similarity 57.2%;
Matches 362; Conservative 0; Mismatches 266; Indels 5; Gaps 2;

Qy   447 GGAGTCCCCTTCTCTCGCGGCGCTTTACAGATGGCCCCAACGAGCCGGTAAGAGGGGACCAAGTTCTTC 506
         ||||| |||||| || ||| | || | ||||| |||| | ||| | ||| | |||||| |
Db   448 GGCCTCGGCGGAGGTCCGCGCTTCCGCTTCGTCCGACGCGCCGACCGGGGGGGCCTCAAGTTCTCC 507

Qy   507 AATGGCGTCCCTGCGGCGCCTCCTTCCCTTGGCCCACGTCCGGTTCCACATTCAACCAA 566
         || |||| ||| ||| | | ||| | || | | || ||| |||| ||||
Db   508 GGCGGCGGCGCACCGTGCGGCGTCGGACGACCCTGCTCGTCCGGTCCGTGGAGCGAG 567

Qy   567 ACTCTGCTCGGAAGAGGCAGGTAAAATGATGGGGCAAAGAGGCCACACGCGTAAGAGTGCGCAT 626
         | |||||||| |||| | | | | || || ||| ||| | |||||||| ||| ||
Db   568 GAGTCGACGACGGAGGTCGGCGGGCTGCTCGCCGGAGGAGCCCTCGGCGCAGCAGATCCAC 627

Qy   627 GTGATCCTCGGCCCGACTATCAACATGCAACGCTCCCCTCGGTGGACGTGGCTTCGAG 686
         || | ||||||||| | |||| | |||| ||| |||| ||| || | | ||||||
Db   628 GTCGTGCTCGGCCCCACGATCAAACCTGCACCGTCCGTGTCCGGCGGACCGCCTGTTCGAG 687

Qy   687 TCGATTGGTGAGGAATCCGTTCCTCGGGGGGCGTTGGGAGCTGCGGGCTTCATCCGAGGCATT 746
         |||| ||| || || || | || | ||| | ||||| |||| | | | | |||
Db   688 GCCTACTCCGAGGAACCCGGTCGCTCACGGCGCCGGGCTCCGCGCTACGTGCGCGGCCCTG 747

Qy   747 CAGGAGCACTGGAGT-GCAGGCGTACGATCAAGCACTTTTTGTGCAATGATGATCAGGAGGACAG 805
         ||||||| | |||| | |||||| |||| | ||| | | |||| | || |||||| |||
Db   748 CAGGAGCCTTCGGCGTCGGGCGGTCGGCGGCCACCTTCGTCAAGGCACTTCGAGGTCGGGAGACCGA 807

Qy   806 GCGCATCGAATGGTTGCAGAGCATCGTCACGGAGGACGGGCCTCTCCGTGAAATCTACGCACTCCC 865
         ||||| | || | | |||| | |||||||| ||| || || | |||| | | |||| |||||
Db   808 GCGCAACACCATGAAACTCCGTCGTCGACCCCGCGAGCTCTTACCTCGTCGCC 867

Qy   866 GTTCCCAGATTGCTCGTGTCCGAGAGACTCCCACGCCGGGTTGGGGTCGACGCGGCGTACAATGGCAT 925
         ||||| |||| | |||| |||| | |||| | |||| | | |||| | ||||||||| ||||
Db   868 GTTCCGAGATNCGCCGTCGACGAGGTCCGAACGGTCTCAAGGCCGCCGAGTG 927

Qy   926 CAATGGCGTGTCGTTGCACGGAGGAACCCTAAAAATATCTTGGATGGGATGCGTTCGAAAGGAATG 985
         ||| |||| | | | |||| | | | | || |||| ||| || | | ||||| |
Db   928 CAACGGCGGTCCCCGCGACGGACCACCACTGCTGTCAACGAGGTCCTCAAGGGCCGAGTG 987

Qy   986 GGGTTGGGATGGGCTAATCATGAGCGACTGGTTACGGCGCACTACAGATACCACAGAACGGCGT 1045
         ||| | ||| |||| | |||| | ||||||||| | ||| | | | | | || ||||
Db   988 GGGCTACACCGGCCTCGTCATGTCACCGGCCACTGGTTCGCGACCCGGACCGCGC 1047

Qy 1046 TGTGCGAGGCCTCGAACTCGACGAGCCGGACC 1078
         | |||||||||||||| |||| |||||||
Db 1048 CG---CCCCTCGAACCTGCTGCATGCCCGGGCC 1076

KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory disease; ds;
KW   hypercholesterolaemia; crop protection agent;
XX
OS   Streptomyces venezuelae ATCC15439.
XX
FH   Key   Location/Qualifiers
FH
FT   CDS   1..2340
FT   /*tag= a
FT   /product= "DesR"
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999; 99WO-US14398.
XX
PR   26-JUN-1998; 98US-0105537.
XX
PA   (MINU) UNIV MINNESOTA.
XX
PI   Sherman DH, Liu H, Xue Y, Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77189.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Claim 3; Page 369-370; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryc gene cluster of Saccharopolyspora erythraea or.
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the
CC   production of biologically active macrolides. The macrolide biosynthetic
CC   proteins are useful for synthesis of methymycin, pikromycin,
CC   neomethymycin and narbomycin. The alternative termination of polyketide
CC   synthesis may be useful to prepare novel antibiotics and
CC   polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC   recombinant host cells are useful as biopolymers, e.g., in packaging or
CC   biomedical applications, to engineer PHA monomer synthases or to prepare
CC   biologically active agents, such as chemotherapeutics,
CC   immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC   disease as well as other diseases involving respiratory inflammation,
CC   cholesterol-lowering agents or macrolide-based antibiotics which are
CC   active against a variety of organisms, e.g., bacteria, including
CC   multi-drug resistant pneumococci and other respiratory pathogens, as well
CC   as viral parasitic pathogens, or as crop protection agents (e.g.,
CC   fungicides or insecticides) via expression of polyketides in plants.
CC   Sequences AAZ87286-287294 represent desosamine biosynthetic genes from
CC   Streptomyces venezuelae ATCC 15439, which encode proteins
CC   AAY77181-Y77189.
XX
SQ   Sequence 2430 BP; 878 A; 878 C; 834 G; 311 T; 0 other;

Query Match   5.9%; Score 175.2; DB 21; Length 2430;
Best Local Similarity 47.6%;
Matches 1016; Conservative 0; Mismatches 968; Indels 150; Gaps 11;

Qy   544 CGCTCGGTTCCACATTCAACCAAACTCTGCCTCGAAGAGGCAGGTAAGATCANGGGCAAAG 603
         | | | ||| |||| | | ||| ||| || || | ||| || | || || |
Db   353 CCCTGGCCAGCACCTTCGACGACCACCATGGCCCACAGCTACGGCGAAGGTCATGGGGCCGG 412

Qy   604 AGGCCATCGCTAAGGAGTGCGGCCATGTGATCCTCGGCCCCGACTATCATCAACATGCAACGCTCC 663
         || || || || |||||| ||||| | ||||||||||| ||| | | | |||||| |||| ||

RESULT 6

AAZ87294
AAZ87294;
ID   AAZ87294 standard; DNA; 2430 BP.
XX
AC   AAZ87294;
XX
DT   05-JUN-2000 (first entry)
XX
DE   S. venezuelae macrolide beta-glycosidase gene desR, SEQ ID NO:23.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;

```
Db   413  ACGGTCCGGCGCGGTTCAACCAGCAGCATGTCTGGGCCCGATGATGAACACATCCGGGTGC                                  472

Qy   664  CTCTCGGTGGGACGTGGCTTGCCTTCCGAGTCGGATTGGTGAGGAATCGGGTTCCTTGGCGGGGCTTTGGGAG                        723
          ||||               || |||                      ||||
Db   473  CGCACGGCGGCGGACTACCAGACCTTCAGCGGAGTCGGATTGGTGAGGAATCCGTTCCTGGCGGGGCACCG                          532

Qy   724  CTTGCGCGGCTTCTCATTCGCGGCGATTCAGAGCACTGTTGCAGGCTACGATCAAGCACTTTT                                  783
          |||| ||||| || |                                  |||||||| |||
Db   533  CGGTCCCCGACGATCAAGGGCATCAAGGGTTCTGATGACCACGCAAGCACTTCG                                           592

Qy   784  TGTGCAATGATGAGGACGACAGGCGCCATGAGTGGTTCACGGGAGAGGGCTC                                             843
          |||  ||||||                                       ||||
Db   593  CGGCCAACAACCAGGAGAAACAACCCGCTTCTCCGGCCAAGTTCGACGACAGAGGC                                         652

Qy   844  TCCGTGAAAATCTACGGCCGACTCCCGTTCCAGATTGCTGTGCGGGACTCCCAGGCCGGTGCCT                                 903
          |||||                                                      ||||
Db   653  TCCGCGAGGATCGAGTTCCCGGCCGTTGCAG---GCTTCCTTCCAAGGCGCGCGGCCCTTCT                                   709

Qy   904  TCATGACGGGCGTACACATCGAATGGGCCATGCGGTGTGCAGCGGCTCGAGAGAACCCTAAATATCTTG                            963
          |||||                                                             ||||
Db   710  TCAATGTGTTGCCTACTACAACGGGCCTCAACGGGAAAGCCGTCTTCTGCGGCAAACGACGAGCTCCTCA                           769

Qy   964  ATGGGATGCTTCGAAAGGAAATGGGGGTTGGGGATGGCCTAANTCATGAGCGACTGGTACGGCA                                 1023
          |||                                                              |||
Db   770  ACAACGTCGCTGCCACGCAGTGGGCGGCTTCCCAGGCCGTTGGAGTGTCCGGACTGGCTCGGCCA                                829

Qy  1024  CATACAGTACCACAGAAGCCGTTGTGGGCAGGCCTCGACGTCGAGATGCCCGGGACCTCCAC                                   1083
          |||||||                                                        |||||
Db   830  C---CCGGGCCACCGGACCGCCATCACCAAGGGCCTCGACCAGGAGATG-                                               873

Qy  1084  GCTTCCGAGGAGAAAACACTCAAGTTCAACGTCTTCCACACGGGAAAGCCCTTTATCCACGTCA                                 1143
          ||                                                             ||
Db   874  -------GGCGTCGAGCTCCCCGGCGGAGTCCCCGAAGGGCACCAGGAGATCT                                            925

Qy  1144  TTGACCAGAGGGCGTAGGGAAGTTCTTCAGTTCGTTCAAGAGTGTGCTTGCCTCCTCCGGAGTGA                                1203
          ||||                                                           |||
Db   926  AGTTCTTCGGCGA---GGGCGTGAGAGACGGCCGTCCGTCCTGAACGGCACGGTGCTCGGCCA                                  982

Qy  1204  CGGAGAACGGCCCCGAGACGACTGTCAACAACACCCCGAAAACGGCAGCTCTCCTCCGGA                                     1263
          |||||                                                  |||||
Db   983  CCGTGCACCGGGGTGTCGGCGGAGGCGGATCGTCGGCCAGAGAAGTTCGGCTCGTTCCTCTG                                   1042

Qy  1264  AGGTTGGCAACGAAGGGCAATCGTGCTGCTGTGAAGAACGAAACGTTTCTGCCCTTGAGCA                                    1323
          ||                                                           |||
Db  1043  CCACTCTCCGGCGCGCGCGCGGCCCGACGAGGCCGCCCAAGGGTGTCCCGCAAGG                                          1102

Qy  1324  AGAGAGAAGAAGAGCTGATTGTCGGGCCCCACACGCACATACCACGGCGGGAG                                            1383
          |||                                                 |||
Db  1103  TCGGCGAGAAACGGCGCGGTGCTCCTGGCCAGGCCACCAGGGGGCTTGGGAG                                             1162

Qy  1384  GCTCTGCGCGCACTCAGGGCCTACTACTACCAGTCACTCCCTTTGACGGCGCAAGCAGC                                      1443
          ||||                                               |||
Db  1163  ACGCCGGGCAAGAGACCATGCGGTCATCGGCGCCGCCCGACGCGTCGAC -CCCAAGGTCACCGGC                                1221

Qy  1444  TCGGACGAGACGGCCGCCATCGTTACACCGTCGCGGCTACCACCGTTCCTCCATTCTTAGGCG                                  1503
          |||                                                            ||||
Db  1222  CTGGGCACCGCGCCCACGTTCGTCCCGGACTCGGACGGGGTTCTTCAACACCATCAAGGCC                                    1281

Qy  1504  AGCAGTGCCTCACGCGCCGACCGGCGCTCGGAGGGGTTCTCAAGCACCCCC                                              1563
          |||                                               |||
Db  1282  CGGCGCGGGTGCGGGGTGGACGTACGAGACGGACCTTCGAGGACATGCAGC                                              1341

Qy  1564  CTGGTACCCCTAACCGCCGCGACACATTGACGAGTCTCTTCACCAAGAAGGGACATGCACC                                    1623
          ||||                                                         |||
Db  1342  ATCCCGGCGGGGAAACCTCAGC-----CGGCGTTCAACC                                                         1375

Qy  1624  TGGTTGGACTACTTACCACCCCAAAGCGGGCAGCAGACACGTGTACGCCCGACATGGAGGGCACGT                               1683
          ||||                                                            ||||
Db  1376  AGGGGCCACCAGCTCGAGCCCGGGCAAGGCGGGCGCTGTACGACGGCTGACCGTGC                                         1435

Qy  1684  ACACCCCCGACGAGGACTGCACCTACGAGCTGCGGCTCGTCTGCGCGGCCAAAGG                                          1743
          |||||                                                 ||||
Db  1436  CCGCCGACGCGGGAGTACCCGCAATCGCGGTCGCCACCGGTGCAG-                                                   1494


Qy  1744  CGTACGTAGACGACCAGCTCGTCGTCGACGGCACGCCACCAAGCAGGTTCCCGGCGGATGCCT                                  1803
          ||                                                          ||||
Db  1495  -----------------------------------CTCGGCGACCACA                                                1507

Qy  1804  TCTTTCGGCTCCCGCCACCCGCGAGGAGAAGACGGGGCCGCATCAATCTCGTCAAGGCCAACACGT                               1863
          ||||                                                             |||
Db  1508  CCATCGAGGCCGGTCAGGTCTTACGCAGGTCTGATGACCGCAAAGGTGAGGACGGGCAACTTCG                                 1567

Qy  1864  ACAAGTTCAAGATGTCAGAGTCGGCCTCCCAACCACCACCTACACCCTCAAGGGGGACCATCG                                  1923
          |||                                                          ||||
Db  1568  GCAACGCACAAGCT---------------CACGAATCGGGCCTTCGGCGATGA                                            1603

Qy  1924  TCCCCGGCCAGGCGTCCCTCCGCGTCGCGAGGCTTCATTGACGAGCCGGGCGAAA                                          1983
          ||||                                                  ||||
Db  1604  GTGGCCACCCCGCCTCCCCCGGAGCGTGGGCTTGGGTGAGCCGCCACCGCAGGA                                           1663

Qy  1984  TCGAAAAGTCGTCGCCCTCGCCCAAGGAGCAGCACGACCAGGTCATCATCAGAGACGG                                       2043
          ||||                                                    ||
Db  1664  TCGCGAAGGCCGTGGAGTCGGCGCGGAAGGCCGTTACGGCGGTCGCCTCGCA------                                       1715

Qy  2044  ACGCCCGACTGGGAGACGGGCAACTCAATTGCGACTCGTCTTTTGGCGGTACNAACCCTTCGG                                  2103
          ||                                                          ||||
Db  1716  -CTACGACGAGGCGCCACCGAGGCGGTGTCCGACAGGCTGGTCCAGAGACGTCACCGAGAGG                                   1774

Qy  2104  ACCAGTCATTGCCGACGTGGCCGCCGAGGGCGTCGACCGTCACCGGACTGGCTTGGCCA                                      2163
          ||||                                                    |||
Db  1775  ACAAGCTGATCTCGGCGTGTGCGGCGACGCCCAACCGGAACAACGACGATCGGTGGTCCTCAAACACG                              1834

Qy  2164  GCAACCCCGAGGAGGAGATGCCCTGGCTCGACGCCACGGCCGCCGGGCCGCTCGGTTACG                                     2223
          ||||                                                      |||
Db  1835  GTTCGTGCCGTGCTGATTGCCGTGGCGGCTCCTTCCTTACGGCAGCCCCAGGCGTGCTGGGTTACC                                1894

Qy  2224  GCGGCAACGAAGACGGGCGAACTCATTGCGCGACTTCGTCTTTTGCCGGGTAC                                            2283
          ||||                                                 |||
Db  1895  CGGGCGGCCCGGCGGGGCGCGCCACCGGCGTCGAGGCGGTCAGGCCACCGGTCCCGGGTTACG                                  1954

Qy  2284  GCAAGCTGTCCCTCAGCGTTCCCC-------------AAGGCGCTGCAGGACGAACCCCGGTTTC                                 2334
          ||||                                                            |||
Db  1955  GCAAGCTCACGCACGAGCTTCCCGGCGGGCCGGAGAACCAGCACGGTGCGGCGGGACCCGA                                    2014

Qy  2335  TCNACTTCGCCACCCGAGGGCCGGCGCCACGCGTTACGGCGGGGTCGGGGTACA                                           2394
          ||                                                   ||||
Db  2015  CAAGCTACCCCGGGGGTCGACAACCAGCAGGACGTACCGCGACTACCACCTCGGGTACC                                      2074

Qy  2395  GGTACTACCGAGTTTGCCGACAAGGACGTCAATTTCCCCGTTCTCACACGTGTCCTACA                                      2454
          ||||                                                     ||||
Db  2075  GCTGGTTCGACAAGGAGAACGTCAAGCCCGTCGTTCCCGGTTCACGGCGGCCGCTGTCTCGTACA                                2134

Qy  2455  CCACTTTTGCCTTTTCCAAATCTCTCCGTTGTCTCACAGGAC-----------GGCAAGCTGAGCGTGT                            2511
          ||||                                                              ||||
Db  2135  CCTCGTTCACGCAGAGAGCGCGCCCCGACCGTCAGTCACGGCTCGAGGGTCA                                             2194

Qy  2512  CCCTCTCCGTGAAGAAACACCCGGTCGTTGGCCGGGCCGGGGGGGCCGGGCTTCAGGTGGCTCACGGCGTCA                          2571
          ||||                                                                  ||||
Db  2195  CGGTCACGGTCCCCAACAGCCGGGGCGCGGGCCAGGAGGTCGTTCCAGGCGTACCTCG                                       2254

Qy  2572  AGCCCCTCCAAGCGGGCAAGATTAACCGCCCGGTTCAAGAGACTCAAGGGCTCCGCAAAGG                                    2631
          ||||                                                           ||
Db  2255  GTGCCCACGCCCGAAACGTGACGGCTCCCAGGGCGAAGAGAAGCTCGTGGGGCTACACGAAGG                                  2314

Qy  2632  TCGAAACTTGCAGCCCGGCGGGGGAGAGGAGAGCGGTGAC                                                         2665
          ||||                                   |||
Db  2315  TCTCGCTCGCGCGGGGGGAGGCGGAGGCGAAGACGGTGAC                                                         2348
```

```
XX
DE   Streptomyces venezuelae DesR gene.
XX
KW   Glycosylated polyketide; modified recombinant bacterial host cell;
KW   mRBHC; macrolide; anthracycline; angucycline; avermectin; milbemycin;
KW   tetracycline; polyene; polyether; ansamycin; isochromanequinone; sugar;
KW   desosamine; DesR; glucosidase; enzyme; gene; ds.
XX
OS   Streptomyces venezuelae.
XX
FH   Key             Location/Qualifiers
FH
FT   CDS             1..2430
FT                   /*tag= a
FT                   /product= "streptomyces venezuelae DesR protein"
FT                   /transl_except= (pos:1..3, aa:Met)
FT                   /note= "CDS does not include start codon"
FT                   /partial
XX
PN   WO20020229035-A2.
XX
PD   11-APR-2002.
XX
PF   05-OCT-2001; 2001WO-US31255.
XX
PR   05-OCT-2000; 2000US-238185P.
XX
PA   (MINU ) UNIV MINNESOTA.
PA   (LIUH/) LIU H.
PA   (SHER/) SHERMAN D H.
PA   (ZHAO/) ZHAO L.
XX
PI   Liu H, Sherman DH, Zhao L;
XX
DR   WPI; 2002-405171/43.
DR   P-PSDB; AAE24237.
XX
PT   Modified recombinant bacterial host cells in which the expression and
PT   activity of nucleic acids encoding sugar biosynthetic enzymes has been
PT   altered, useful for producing metabolites with altered sugar structures
PT   -
XX
PS   Disclosure; Page 170-171; 174pp; English.
XX
CC   The invention provides a method to alter the sugar structure diversity
CC   for a particular metabolite via the recruitment and collaborative action
CC   of sugar genes from a variety of sugar biosynthetic pathways to yield a
CC   polyketide. The invention also relates to a modified recombinant
CC   metabolite comprising a non-natural sugar, e.g., a novel glycosylated
CC   bacterial host cell (mRBHC) in which the expression and activity of
CC   nucleic acids encoding sugar biosynthetic enzymes has been altered.
CC   The mRBHCs may be cultured to produce the modified sugar products,
CC   e.g. a macrolide, anthracycline, angucycline, avermectin, milbemycin,
CC   tetracycline, polyene, polyether, ansamycin or isochromanequinone.
CC   The present sequence is Streptomyces venezuelae sugar (desosamine)
CC   biosynthetic gene cluster DesR (glucosidase) gene.
XX
SQ   Sequence 2430 BP; 407 A; 878 C; 834 G; 311 T; 0 other;
XX

Query Match             5.9%;  Score 175.2;  DB 24;  Length 2430;
Best Local Similarity   47.6%;
Matches 1016;  Conservative 0;  Mismatches 968;  Indels 150;  Gaps 11;
                              Pred. No. 2e-19;
```

```
Qy   544  CGCTCGGTTCCACATTCACCAAAACTCTGCTCGAAGAGCAGTAAGATGGGCAAAG  603
Db   353  CCCTGGCCAGCACCTTCGACGACACCATGGCCAGCCTACGGCCCGG  412

Qy   604  AGGCCATCGCTAAGAGTGCGCCATGTGATCCTCGGCCCGACTATCAACATGCAACGCTCC  663
Db   413  ACGGTCGCGCGCTCAACCAGGACACATGGTCTCTGGGCCCGATGATGAACAACATCCGGGTGC  472

Qy   664  CTCTCGGTGGACGTGGCGTTCGAGTCGATTGGTGAGGATCCGTTCCTCGGCGGGCTTGGGAG  723
Db   473  CGCACGGCGGCGGCGGGACTACGAGACCTTCAGGAGGACACCCCTGGTCTCCTCGGGCACCG  532

Qy   724  CTGCGGCGTCTCATCCGGCCGCATTCAGAGCACTGGAGTGCAGGCTACGGATCAGCGCACTTTT  783
Db   533  CGGTCGCCCCAGATCAAAGGGCATCAAGGGGTGCGGGTTCTGATGACCACGGCCAAGCGACTTCG  592

Qy   784  TGTGCAATGATCAGGAGGACAGGCGCCGCATAGGTGGTGCGGAGAGCATCGTCACGGGGGTC  843
Db   593  CGGCCAACAACCAGGAGAACAACCGCCGTTCTCCGGTGAACGGCCAAATGTCGACGACGAGACGC  652

Qy   844  TCCGTGAAAATCTACGCCACTCCGTTCCGATTGCTTGTGTGCGGAGATTGCTTCCCAGCCTGGGTGCGT  903
Db   653  TCCGCGAGATCGAGTTCCCGGCGTTGCAG---GCGTCCTCCAAGGCCGCGGGGCGCCTCCT  709

Qy   904  TCATTGACGGGCGTTACAATGGCCATCAATGGCGTTGTCGTTGCACGCGAGAACCCTAAATATCTTG  963
Db   710  TCATCGTGTTGCCTTACACAACGGGCCTCACGGGGAAGAACCCGCTTCTCGGTCCGACGACGAGACG  769

Qy   964  ATGGGAGATGCTTCGAAAGGAATAGGGGTTGGGATGGCCTAATCATGAGCGGACTGGTTACGGCA  1023
Db   770  ACAACGGTGCTGCGGCACCGCAGTGGGGCTTCCAGGGCTGTGGGTTGCGATTGGTTCCGGCCA  829

Qy  1024  CAATACAGTACCACAGAGAAGCCGTTGTGTGGCAGGCCTCGACCTCGAGATGCCCGGGACCTCCAC  1083
Db   830  C---CCCGGGCCACCGACGCCATCACCAAAGGGCCTTCGACCAGGAGATG-------------  873

Qy  1084  GCTTCCGGAGGAGAAACACTCAAGTTCAACGTCTCCAACGGAAAGCCCTTTATCCCACGTCA  1143
Db   874  ---------GGCGTCGAGGTCCCCGGCCGACGTCCCGGCAACCGGGCGCGCGGCCA  925

Qy  1144  TTGACCAGAGGGCTAGGGAAGTTCTTCAGTTGTCTGTCAAGAAGTGTGTGCTGCCCCGGAGTGA  1203
Db   926  AGTTGTCTTCGGCGA---GGCGCTGAAAGACGACGCCGTCCTTGAAACGGCACGGGTCCCCGAGGCGG  982

Qy  1204  CGGAGAACGGGCCCCGAGGACGACTGTCAACACACCCGAACAAGAGCTCTCCTCGGA  1263
Db   983  CCGGTGACGCGGTGGGCGGAGCGGGGATCGTCGCCCAGATGAGCGGGTCGGTTCGCTCTCG  1042

Qy  1264  AGGTTGGCAACGAGGGGCATGGCGTTGCTGATGAAGAACGAGAACAACAACGTTCGTGCCCTTGAGCA  1323
Db  1043  CCACTCCGGCGGCCCGGCGCCCGAGCGCGACAAAGGCGGGTGTCCCAGGCGCGTGTGCCCCAAGG  1102

Qy  1324  AGAAGAAGAAGAAGACCCTGATTGTGTCGGGCCCCAAGCGCCAAGAGCAGCGGGAG  1383
Db  1103  TCGCCCAGGAACGGCGGCGGGTGCTCCTGCGGCCAGGCGCCCTGCCGTCGGCCGGTG  1162

Qy  1384  GCTTCTGCCCGCACTCAGGGGCCTACTACGGCAGTCACTCCCCTTTGACGGGCCTCAGCAGCAGC  1443
Db  1163  ACGCCCGGCGAAGAGCATCGGCGGGTCATCGGGCCCGTCATCGGGCTCGAC-CCCAAGGGTCACCGGC  1221

Qy  1444  TCGAGACGACCGCGCCATCGTACACCGTTCGCGGGCCCTACACCACCGTTGTCCCCATTCTTAGGCG  1503
Db  1222  CTGGGCAGCCGCCCGCGACGCGGGGCGGGGCCCGACTCGGCCCACCACCACCATCAAGGCC  1281

Qy  1504  AGCAGTCGCCTCACCGCCCACGTTCCGGGGTCCCGGACTCGGGCGGGTGTCTTTCAACCAGCCC  1563
Db  1282  CGCGCGGGGTCGGGGTGCGACGTACGAGACGGGGTGAGGGAGACCTTCGGGGACCGGAG  1341

Qy  1564  CTGGTACCCCTAAACCGCGACGACATTGACGAGGTCTCTTCACCAAGACGGACAATGCACC  1623
Db  1342  ATCCCGGCCGGGGGGAACCTCAGC---------------CCGGCGGTTCAACC  1375

Qy  1624  TGGTGGACTACTTACCACCCCAAAGGCGGCAGAACACACGTGGGTACGACGACCATGGACAGT  1683
Db  1376  AGGGCCACCAGCTCGAGCCGGGGCAAGGCGGGGCCCGGTGTACGACGGCACGGGCGACCAGGAAGG  1435

Qy  1684  ACACCCGCCGACGAGGACTGCACCACGGCCTCGGCCTTCGTCGTTGGGGCGGGGCAAAGG  1743
Db  1436  CGCCCGACGGCGGCGAGTACCGCATCGCGCGGGTCCTGCCGGTGGTTGCCGCCAACGCGGGAGGTGC  1494

Qy  1744  CGTACGTAGACGGACCAGCTCGTTCGTCGGACAACGCACCAAAGCAGGTGCCCGGCGGAATGCCT  1803
Db  1495  -------------------------------------CTCGGGCAGGCCACA  1507
```

```
Qy   1804 TCTTCGGCTCCTCCGCCACCCGGAGGAGGACGGGCGCGATCATCTGTCAAGGGCAACACGT 1863
               | |||  ||  |||
Db   1508 CCATCGAGGCCGCGGTCAGGTCTACGGCAAGGTGAGCAGCGCGCCTCAAGCTGACCAAGG 1567

Qy   1864 ACAAGTTCAAGATCGACGAGTTCGGCTCCGCACCACTTACACCTCAAAGGGCGACACCATG 1923
               ||  || |||  ||||     ||| |||| |||
Db   1568 GCACGCACAAGCT------------------------CACGATCTCGGCGCTTCCGCGATGA 1603

Qy   1924 TCCCCGGCCCACGGGCTCCCTCCGCGGTCGGCGGCGTCGCAAGGTCATTGACGACCAGGCGCGAAA 1983
               ||||| ||| |||
Db   1604 GTGGCCACCCCGCTCTCCTGGAGCGTGGGCGTCGGCTGCGTGGGTGAGCTGGGC 1663

Qy   1984 TCGAAAAGTCGGTCGCCGCTTCGCCCAAGAGGAGCACGACCAGGTGCATCATCTGCGCGGGGCCTTA 2043
               ||| ||| | ||| |||||
Db   1664 TCGCGAAGGCCGTGAGTCGGCGCGGAAGGCCCGTACGCCGGGTCGTCTTCGC-------- 1715

Qy   2044 ACGGCCGGACTGGGGAGGACGCGAGGGGCGCGAGCATGAAGCTCCCCGGCGGGTGCTGG 2103
               || | ||| | |||| |||
Db   1716 -CTACGACGGACGCACCGAGGGGGTTCGAACGCGTGGTGGGCGGGTACGGACGGA 1774

Qy   2104 ACCAGCTCATTGCCGACGTGGGCCCGGCCGGCGGCGGGGTGCTGGTCGTCAGCAGCACGG 2163
               || || ||| ||||| | ||| ||| | |||
Db   1775 ACAAGCTGATTCTCGGCTGTCGCGACGCCAACCCGAACGACAGATCGGTGCTTCCTCAACACG 1834

Qy   2164 GCACCCCCGAGGAGATGCCCTGGCTCGACGACCGGCCACACCCGTCATCCGGGTGGTACG 2223
               | ||| | |||| | ||| |||||| | || ||
Db   1835 GTTCGTTCGGGTGCTGATTGCCGTGGCTGTTCCTCAAGACCCGCGGTGCCGGTGGTACC 1894

Qy   2224 GCGGCAACGAGACGGGGCAACTCCATTGCCGACGTCGTCTTTGGCGACTACACCCCTCGG 2283
               | ||| ||||||  |||||| | || |||
Db   1895 CGGGCGCCAGCGCGGGGGCCCGAGGCCGGCCCGTCTTCGCCGGGTAGGTCACCCGAGCG 1954

Qy   2284 GCAAGCTGTCCCTCAGCTTCCC---------AAGCCCCTGCGAGGACAAACCCCCGGTTTC 2334
               ||||| ||
Db   1955 GCAAGCTCACGCAGCAGAGCTTCCCGGCGCGAGAAACCACGCGGTGCGCGACGACCCGA 2014

Qy   2335 TCAACTTCCGCACCGAGGCGCGGCGCACGCGTACGCCACGGTCTACGTCGGGTACA 2394
               ||||| |||||| ||| ||
Db   2015 CAAGCTACCGGGGCGTCGACAACGACGCGAGCGGCCATCCACGTCGGGTACC 2074

Qy   2395 GGTACTAGGAGTTTGCCGACAAGGAGGACGTCAATTCCCTTTGGCGACGGCCGTGTCCTTACA 2454
               || ||| ||| |
Db   2075 GCTGGTTCGACAAGGAGGAACGTCAAGCCGCGTTCCGGTTCGGGCACGGCCGTGTGGTACA 2134

Qy   2455 CCACTTTTGCCTTTTTCCAATCTCTCACAAGGAC---GGCAAGTGAGCGTGT 2511
               || ||| |||
Db   2135 CCTTCGTCACGCGAGAGCCCCGACCGTCGTACGTCCACGGGTGGTCTGAAGGTCA 2194

Qy   2512 CCCTCGTCCGTGAAGAACACCGGGCTCTCCGTGCCCGGCCAGCGTCGTACGTCA 2571
               |||| |||
Db   2195 CGGTCACGGTCGCGACAACGGGGAAGGTCGTCAGGGCGTACCTCG 2254

Qy   2572 AGCCCGTCCAAGCGGCCAAGATTAACCGCCCGTCAAGAGACGCTTCGCAAAGG 2631
               || |||
Db   2255 GTGCGACGGCCGGAACGTGACGAGCGGCTCCCGCCCGGAGCGTACGATCGTGGGCTACGAACGG 2314

Qy   2632 TCGAACTGCAGCCCGGCGACGAAGGCGGTGAC 2665
                  ||| |||  || |||||||||||
Db   2315 TCTCGCGTTCGCCCGGCGCGGGCGCGGAGCGACGTGAC 2348
```

RESULT 8

```
AA287284
ID   AA287284 standard; DNA; 12441 BP.
XX
AC   AA287284;
XX
DT   05-JUN-2000 (first entry)
XX
DE   S. venezuelae desosamine biosynthetic gene cluster plkB, SEQ ID NO:3.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent; ds.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH, Liu H, Xue Y, Zhao L;
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Claim 2; Page 281-287; 438pp; English.
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77179.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryc gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the production
CC   of biologically active macrolides. The macrolide biosynthetic proteins
CC   are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC   narbomycin. The alternative termination of polyketide synthesis may be
CC   useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC   monomers. The compounds produced by the recombinant host cells are useful
CC   as biopolymers, e.g., in packaging or biomedical applications, to
CC   engineer PHA monomer synthases or to prepare biologically active agents,
CC   such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC   chronic obstructive pulmonary disease as well as other diseases involving
CC   respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC   antibiotics which are active against a variety of organisms, e.g.,
CC   bacteria, including multi-drug resistant pneumococci and other
CC   respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC   protection agents (e.g., fungicides or insecticides) via expression of
CC   polyketides in plants. The present sequence represents the desosamine
CC   biosynthetic gene cluster from Streptomyces venezuelae ATCC 15439.
XX
SQ   Sequence 12441 BP; 4294 A; 4686 C; 1757 T; 0 other;
```

```
Query Match             5.9%;  Score 175.2;  DB 21;  Length 12441;
Best Local Similarity   47.6%;
Matches  1016;  Conservative    0;  Mismatches  968;  Indels  150;  Gaps  11;
```

```
Qy    544 CGCTCGGTTCACATTCAACCAAACTCTGCTCGAAGAGGCAGGTAAGATGAATGGGCAAAG 603
              | |||          ||||       |||          |||
Db   3977 CCCTGGCCGAGCAGCACCTTCGCACGCACCACCATGGCCGACAAGCTACGGCAAGGTCATGGGCGCCCG 4036

Qy    604 AGGCCAATCGCTAAGAAGTGCCGCATGTGATCCGTCGCCCGACTATCAACATGGCAACGCTCCC 663
              |||          |||                |||||||       |||     |||
Db   4037 ACGGTGCGGCGCCCTCAACCACGAGCACATGGGTCCTGGGCCCCGATGATGAACATCCGGGTGC 4096

Qy    664 CTCTCGGTGGACGTGGCGTTCGAGTGGTTGTTGGTGAGGATCCGTTCCTCGCGGGCTTTGGGAG 723
              ||||              |||       |||           |||||||||   |||
Db   4097 CGCACGGCGGCCGGAACTAGGAGACCTTCAGCGAAGCACTGGAGTCGAGGAGGACCCCCTGGGTTCCTCGCGGCTACGATCAAGCACTTTT 4156

Qy    724 CTGCGGCTCTCATCCCGCGGGCGATTCCGGGGCAGCGAGCACTTTT 783
              ||                 |||           ||| |||||||||
```

```
Db  4157 CGGTCGCCCCAGATCAAGGGCATCCAGGCTGCTCGATGACCACGCGGCCAAGCACTTCG 4216
Qy   784 TGTGCAATGATCAGGAGGACAGCGGCGCATGAGTGGTGCAGAGCATCGTCACGGAGCGGCTC 843
              ||||  ||||   |||  |||  || |||  ||| |||| ||  |||  ||| |||
Db  4217 CGGCCAACAACCAGGAGAACAACGCGCTTCTCCCGTGAACGCAATGTCGACGAGCAGAGC 4276
Qy   844 TCCGTGAAATCTACGCACTCCCGTTCAGATTGCTGTGCGAGACTCCCAGCCGCGGTCCGT 903
              |||  ||| |||||   ||| ||||  ||| |||  ||| |||   |||  |||
Db  4277 TCCGCGAGATCGAGTTCCCGGCGTTCGAG---GCGTCCTCCAAGGCCGGGCGGGCCTCT 4333
Qy   904 TCATGACGGCGGTACAATGGCATCAAATGGCGTGTCGCGGAGAGAACCCTAAATATCTTG 963
              || ||| |||| |||  ||  || |||  |||  || |||  ||  |||
Db  4334 TCATGTGTTGCCTACAACGGACGTCAAACGGGGAAGCCGTCCTGCGCCAACGACGAGCTCCTTCA 4393
Qy   964 ATGGGATGCTTTCGAAAAGGAATGGGTTGGGGATGGCCTAATCATGACGGCACTGGTTACGGCA 1023
              |||  ||| |||||  |||  |||  ||| || ||| |||  |||  |||  |||
Db  4394 ACAACGTGCTTGCCGCCACGGTGGGGCGTTCCAGGGCTGATGTCCGACTGCTCGGCCA 4453
Qy  1024 CATACAGTACCACAGAGACAGCCGTTGTTGGGCAGGCTTCGACCTCAGATGCCAGGAACCTCCAC 1083
              |||  ||| |||  |||  ||| |||  ||| |||  ||| |||   |||
Db  4454 C---CCCGGGCACCGACCGCCATCACCAAGGGCCTTCGACCAGGAGATG------ 4497
Qy  1084 GCTTCCGAGGGAGGAGAAACACTCAAGTTCAACGTCTTCCAACGGAAAAGCCCTTTATCCACGTCA 1143
              |||  |||  |||  |||  ||| ||| ||| ||  ||| |||  ||| |||
Db  4498 ------GGCGTCGAGCTTCCCGGCGACGTCCGGAGCCGTTCGCGCGGCCA 4549
Qy  1144 TTGACCAGAGGGCTAGGGAAGTTCTTCAGTTCGTTCAAGAAGTGTGCTGCCTCGGGAGTGA 1203
              || ||| |||  |||   ||| |||  ||| |||  ||| |||  |||  |||
Db  4550 AGTTCTTCGGCGA---GGGCGTGAAGAGGACGCACTCTCCTCCGGA 4606
Qy  1204 CGGAGAACGGGCCCCGAGAGACGTGTCACACAACCACCCGGAAACGCGACTCTCCTCCGGA 1263
              ||||   |||  ||||  ||  ||| |||  |||  ||| |||  |||
Db  4607 CCGTGACAACGCGGTCGCGGGAGGATCGTTCCCCAGATGGGAGAAGTTCGGTCGTCCTCG 4666
Qy  1264 AGGGTTGGCAACGACGGGCATCGTGCTGTCTGAAGAACGAGAGAACAACGTTGTCCCTTGAGCA 1323
              |||  ||| ||| |||  |||  ||| ||| |||  |||  ||| |||  |||
Db  4667 CCACTCCGCGCCCGGCCGCCCGGACCAAGCGGGGGTTGCCCAGGCGGTGTCCGCAAGG 4726
Qy  1324 AGAAGAAGAGAGACGGCTGATTGTCGGCCCCGAACGAGGCCCACATACCACGCGGAG 1383
              ||||  |||  |||  ||| |||  |||  |||  ||| |||  |||  |||
Db  4727 TCGCGCGAGAACGGGCGGTGCTCTGCCGCAAGCGAGGGCCTGCCGCGTG 4786
Qy  1384 GCTCTGCGCCACTCAGGGCCTACTACCGACAGTCACTCCCTTTGACGGCAAAGCAGC 1443
              ||||  ||| |||  |||  ||| |||  |||  ||| |||  |||  |||
Db  4787 ACGCCGGCAAAGAGATACTGCGGTCATCGGCCCGACCGCCGTCGAC-CCCAAGGTCACCGGC 4845
Qy  1444 TCGAGACCGCGCGCCATCGTACACCGTCGGCGGCCTACACCGTTCCTCCCATTCTAGGCG 1503
              ||||  ||| |||  |||  ||| |||  |||  ||| |||  ||| |||
Db  4846 CTGGGCAGCGCCCACGTGTCTCCCGGACTCGGCGGCGCCCACTCGACAGCCATCAAAGGCC 4905
Qy  1504 AGCAGTGCCTCACGGCCCGACGGGCGGCGGCCTCCGGCGGCGTCTTCAACGAGGCCCC 1563
              ||||  |||  |||  |||  ||| |||  |||  ||| |||  |||
Db  4906 CGCGCGGGGTGCGGGGTCGCGGGTGACGTTACGAGACGGGGACACGCAG 4965
Qy  1564 CTGGTACCCCTAACCGGCCACACAATTGACGAGCTCTTCTTCACCAAGGACAC 1623
              ||||  |||  |||  |||  ||| |||  |||  ||| |||
Db  4966 ATCCCGGCGGGGGAACCTCAGC------CCGGCGTTCAACC 4999
Qy  1624 TGGTGGGACTACTACCACCCCAAGGCGGCCAGACAACGGGCACT 1683
              |||  |||  |||  |||  |||  ||| |||  |||
Db  5000 AGGGCCACCAGCTCGACGCGGGCAAGGCGGGGGCCTCGTACGGGGCAGCTCACGGTGAC 5059
Qy  1684 ACACCGCCGACGAGGACTGCACCTACGAGCTCGGCCTCGGTGTTACGGCCAAAGG 1743
              |||  |||  |||  |||  ||| |||  |||  ||| |||
Db  5060 CCGCGGACGGGCGGCGAATCGCGGTCGCCTCGTTGCGCGGATGCCTGGCGCAG- 5118
Qy  1744 CGTACGTAGACGAGCTCGGTCGTCGACAAACGGTCACCGGTGTTCACCGGTGCGCCGC 1803
              |||  ||| |||  |||  ||| |||  |||  ||| |||
Db  5119 ------------------------------CTCGGCAGCCACA 5131
Qy  1804 TCTTCGGCTCCGCGGCGAGGAGGACCCCGGAGGAGAGACGGCCCCCGCCGAGCTCGCCTGAC 1863
              |||  ||  ||||  || |||  |||  |||
Db  5132 CCATCGAGGCCGGTCAGGGTCTACGGCACGGTCCTCAAGCTGACCAAGG 5191
```

```
Qy  1864 ACAAGTTCAAAGATCGAGTTCGGCGTGGTCTCCCGGCCACCTACACCTTCAAGGGCGACACCATCG 1923
              ||||  |||  |||  |||  ||| |||  |||  ||| |||  |||
Db  5192 GCAACGCACAAGCT-------------------CACGATCTCGGCGTTCGCGATGA 5227
Qy  1924 TCCCCGGCCACGGCGTCCTCCGGGCGTGCGGTCGCGCCGAAGGTCATTGACGACCAGCCGAAA 1983
              ||||  |||  |||  |||  ||| |||  |||  ||| |||  |||
Db  5228 GTGCCACCCGCGTCTCCCTGGGACGTGGTCTGGGTGACGCGCACGCGACGA 5287
Qy  1984 TCGAAAAGTCGTGCGTCGCCCTCGCCAAGGACGACACAGTCATCATCTGCGGGGGCCTTA 2043
              |||  |||  ||  ||||  ||  |||  |||  |||  |||  |||
Db  5288 TCGGCGAAGGCGGTGGAGTCGGCGGGGAGGCCCGGTAGCGGGGTCG------- 5339
Qy  2044 ACGGCGACTGGGGAGGACGCCGAGGGGCGACCGGCGGCATGAAGCTCCCGGCGCGTGCGG 2103
              |||  |||  |||  |||  ||| |||  |||  ||| |||  |||
Db  5340 -CTACGACGGACGGCGACCCGGAGGGGGTCGGACGGTCCGACACTGGGCGTCCGGGTACC 5398
Qy  2104 ACCAGCTCATTGCCGACGTGGCCCCGGCCCGGCGGGGGGCGAACCCAAACACCGTGCTCAGGACGG 2163
              |||  ||| |||  |||  ||| |||  |||  ||| |||
Db  5399 ACAAGCTGATCTCGGCGTTCCCGGCGGCCAACCCGAACAGAGTTGTCCG---GTGGTGGGTGTTTCA 5458
Qy  2164 GCACCCCCGAGGAGAGAATGCCCTGGCTGACGCCACGCGCCGTGTCATCCAGGCCTGGTACG 2223
              ||||  |||  |||  |||  ||| |||  |||  |||  |||
Db  5459 GTTCGTCGGTGCTCGATGCCGTGGCTGTCCAAGAACCGTCCTGCGCGACCTGTCGGTGTACA 5518
Qy  2224 GCGGCAACGAGAGACGGGCCAAACTCCATTGCCGACGTCGTCGTTTGCGGCTACTRACACCCCTCGG 2283
              ||||  |||  ||| |||  ||| |||  |||  ||| |||  |||
Db  5519 CGGGCCAGGCGGCGGCCCCGAGGGTGGCCACCGGTGCCTGCTTGTGACTGCAACCCGGCG 5578
Qy  2284 GCAAGCTGTGCCCTCAGCTTCCC------------AAGGCGCCTGCGAGGACAACCCGCGTTTC 2334
              ||||  |||  |||  |||  ||| |||  |||
Db  5579 GCAAGCTCACGACAGCAGTTCCCGCGCCGCGGGCCAGGCCAACCCGACGGTCG 5638
Qy  2335 TCAACTTCGGCACCGAGGCGGGGGCGGGCGCGGGTGGACTACGTGCGGCGAAGG 2394
              ||||  |||  |||  |||  |||  ||| |||
Db  5639 CAAGCTACCCGGCGTCGACAAGGACGGTCAAATTTCCCGGTACGGCGGAG 5698
Qy  2395 GGTACTACGAGTTTGCCGGACAAGAGAAGACGGTCAAGACGTACCGGCGACCCGA 2454
              ||||  |||  |||  |||  ||| |||  |||  |||
Db  5699 GCTGGTTCGACAAGGAGAAGTCAAGCGCCTGTTCCGGTTCGGCACGGCGTCGTACA 5758
Qy  2455 CCACTTTTTGCCTTTCAAATCTCTCCGGTGTCTCACACAGGAC---GGCAAGCTGAGCGTGT 2511
              ||||  |||  |||  |||  |||  ||| |||
Db  5759 CCTCGTTCACGCAGCAGGGCCCCGACCGTCGTGGTCCACCGGTTCGCAGAGGTCA 5818
Qy  2512 CCCTCTCCGTGAAGAAACACCGGGTCCTCCGGGCGCACACGGTGGCCCAGCTCTACGTCA 2571
              ||||  |||  |||  |||  ||| |||  |||
Db  5819 CGGTCACGGTCCCGAACACGCGGAAGCGCGGCACGCGAGGGTCGTCAGGCGGTTACCTCG 5878
Qy  2572 AGCCCCTCCAAAGCGGCCAAGATTAAACCGCGCCCCGTCAAGGAGGCTCAAGGGGCTTCCCGAAAG 2631
              ||||  |||  |||  |||  ||| |||  |||  |||
Db  5879 GTGCGCAGCCCGAACGTGACGGGCTCCAGGGCACACGGCGCGAAGAAGTCGTGGGCTACACGAAGG 5938
Qy  2632 TCGAACTGCAGCCCGGCGGAGCGAGACGAAGGCGGTGAC 2665
              ||||  |||  |||  |||  ||| |||
Db  5939 TCTCCGCTCGCGCGCGGGCGGAGGCGAAGACGGTGAC 5972
```

```
OS   Streptomyces venezuelae.
XX
FH   Key              Location/Qualifiers
FT   CDS              3..809
FT                    /*tag= a
FT                    /product= "streptomyces venezuelae Des gene cluster
FT                    encoded protein #1"
FT   CDS              806..2014
FT                    /*tag= b
FT                    /product= "Streptomyces venezuelae Des VIII protein"
FT                    /transl_except= (pos:806..808, aa:Met)
FT                    /partial
FT   CDS              2162..3442
FT                    /*tag= c
FT                    /product= "Streptomyces venezuelae Des VI protein"
FT                    /transl_except= (pos:3532..3534, aa:Met)
FT                    /partial
FT   CDS              complement 3535..4245
FT                    /*tag= d
FT                    /product= "Streptomyces venezuelae Des VII protein"
FT                    /note= "CDS does not include start codon"
FT                    /partial
FT   CDS              4312..6741
FT                    /*tag= e
FT                    /product= "Streptomyces venezuelae Des R protein"
FT                    /transl_except= (pos:4312..4314, aa:Met)
FT                    /note= "CDS does not include start codon"
FT                    /partial
FT   CDS              complement (7969..6829)
FT                    /*tag= f
FT                    /product= "Streptomyces venezuelae Des V protein"
FT                    /transl_except= (pos:7969..7967, aa:Met)
FT                    /note= "CDS does not include start codon"
FT                    /partial
FT   CDS              complement (8979..7966)
FT                    /*tag= g
FT                    /product= "Streptomyces venezuelae Des R protein"
FT                    /transl_except= (pos:8979..8977, aa:Met)
FT                    /note= "CDS does not include start codon"
FT                    /partial
FT   CDS              complement (9111..9989)
FT                    /*tag= h
FT                    /product= "Streptomyces venezuelae Des IV protein"
FT                    /partial
FT   CDS              complement (11659..10182)
FT                    /*tag= i
FT                    /product= "Streptomyces venezuelae Des IV protein"
FT                    /partial
FT   CDS              complement (12883..11636)
FT                    /*tag= j
FT                    /product= "Streptomyces venezuelae Des IV protein"
FT                    /transl_except= (pos:12881..12883, aa:Met)
FT                    /note= "CDS does not include start codon"
FT                    /partial
```

```
PT   Modified recombinant bacterial host cells in which the expression and
PT   activity of nucleic acids encoding sugar biosynthetic enzymes has been
PT   altered, useful for producing metabolites with altered sugar structures
XX
PS
PT
XX
XX
PS   Disclosure; Fig 8; 174pp; English.
XX
     The invention provides a method to alter the sugar structure diversity
     for a particular metabolite via the recruitment and collaborative action
     of sugar genes from a variety of sugar biosynthetic pathways to yield a
     metabolite comprising a non-natural sugar, e.g., a novel glycosylated
     polyketide. The invention also relates to a modified recombinant
     bacterial host cell (mRBHC) in which the expression and activity of
     nucleic acids encoding sugar biosynthetic enzymes has been altered.
     The mRBHCs may be cultured to produce the modified sugar products,
     e.g. a macrolide, anthracycline, angucycline, avermectin, milbemycin,
     tetracycline, polyene, polyether, ansamycin or isochromanequinone.
     The present sequence is Streptomyces venezuelae sugar (desosamine)
     biosynthetic gene cluster.
```

```
XX
SQ   Sequence 13613 BP; 1858 A; 4732 C; 5093 G; 1930 T; 0 other;

Query Match             5.9%;  Score 175.2;  DB 24;  Length 13613;
Best Local Similarity  47.6%;  Pred. No. 2.2e-19;
Matches 1016;  Conservative   0;  Mismatches 968;  Indels 150;  Gaps 11;

Qy  544 GGCTCGGTTCCACATTCAACCAAAACTCTGCTCGAAGAGGCAGGTAAGATGATGGGCAAAG 603
Db 4664 CCCTGGCCAGGACCTTGCGACCAGACACCATGGCCGCACAGCCTACGGCGAGTCATGGGCGCG 4723

Qy  604 AGGCCATCGCTAAGAGTGCCGCATGTGATCCTCGGCCCGGACTATCAACATGCAACCGTCCC 663
Db 4724 ACGGTCCGCGCCTCAACCAGACACATGGTCCTNGGGCCCGGATGAGCAACACATCCGGTGC 4783

Qy  664 CTCTCGGTGGAGGTGGCTTCGAGTCGATTGGTGAGGAGTCCGTTCCTGGCGGGCGTTGGGAG 723
Db 4784 CGCACGGCGGCCGGAACTACGAGAACCTTCAGGCAGGAGACCCCCTGGTCTCCTTCGGCACCG 4843

Qy  724 CTGCGGCGCTCTCATCCCGGCAATTCAGAGCACTGGCAGGTGCCAGGCTACGATCAAGCACTTTT 783
Db 4844 CGGTCGCGCCAGATACAAGGGCAATCCAGGGTGCGGGTCAGACGGCCAAGCACTTCG 4903

Qy  784 TGTGCAATGATCAGGAGGACAGGCCCCATGATGGGTGCAGAAGCAGCATCGTCACGGAGCCGGGCTC 843
Db 4904 CGGCCAACAACCAGGAGAACAACCGCTTCTCCGTGAAAGCCAATGTCGACGAGCAGACGC 4963

Qy  844 TCCGTGAAAATCTTACGGCGACTCCCGGTTCGCGGAGACTCCCAGCGGGGTGCGT 903
Db 4964 TCCGGCAGATCGAGAGTTCCCGCGCGTCGAG---GCCGTCCTCCAAAGCGCGGGCGCCCCTCCT 5020

Qy  904 TCATGACGGCGGTACAATGGCATCAATGGCGGTCGTCGTCGCAGGGAGAACCCTAAATATCTTG 963
Db 5021 TCATGTGTGCCTACAACGGGCCTCAACGGGAAGCCGTTCCTCGGCAACGACGAGCTCCTCA 5080

Qy  964 ATGGGATGCTTCGAAAGGAAATGGGGGTTGGGCATGGGGCTAAATCATGACGGACGACTGGTTACGGCA 1023
Db 5081 ACAACGTGCTCGCGCACCGCACCGAGTGGGGCGCTTCCAGGGCGTGGGTGATGTCCGACGACTGGCTCGCCA 5140

Qy 1024 CATACAGTACCACACAGAGACCGTTGTGGCAGGCCTCGACGATGCCCGGACACTCCAC 1083
Db 5141 C---CCCGGGCACCGCAGCCCATCACCCAAAGGGGCTCGACCAGGAGATG------------- 5184

Qy 1084 GCTTCCGGAGGAGAAACACTCAAGTTCAAGTGTCTCCAAACGGAAAGCCCTTTATCCACGTCA 1143
Db 5185 -------GGCGTCGAGCTCCCCGGCAGCGTCCCGAAAGGCGGCCCGAGCGGGGCCA 5236

Qy 1144 TTGACAGAGAGGGCTAGGGAAGTTCTTCAGTTCGTCAAAGAAGTGTGCTGGTCCCGGAGTGA 1203
Db 5237 AGTTCTTCGGCGCA---GGCGCTGAAGACGGCGCCGTCCTGAACGGACGAGCTCCCCGGAGGCGG 5293

Qy 1204 CGGAGAACGGCCCCGAGACGACTGTCAACAACACCCCGAAACGGCCAGCTCTCCTCCGGA 1263
Db 5294 CCGTGACGCGCGTCGCGGACGGATCGTCGGCCAGATGAGGAAGTTCGTTCGTTCCTCG 5353
```

```
Qy   1264 AGGTTGGCAACGCAAGAGGGCATCGTTGCTGAAGAACGAGAACAACGTTCTGCCCTTGAGCA 1323
              |||  || |||||| || | | | | |||| ||| || ||||| | | || ||||||
Db   5354 CCACTCCGGCGCGGCGCGCGAGCGCGACGACGAGGCGGTGCCCAGCGACGTGTCCGCCAAGG 5413

Qy   1324 AGAAGAAGAAGACGCTGATTGTGCGGCCCCAACAGGCGCACATACCACGGGCGGAG 1383
              |||  ||   |||| |||  ||| ||||||   || ||||| | |||||||
Db   5414 TCGGCGAGAACGGCGGCGGTGCTCGGGCAACGCCAGCGCCCTGCGCCTGCGCGGTG 5473

Qy   1384 GCTGCTGCCGCGCACTCAGGGCCTACTACGGCGTCCTTGACGGCGCTCAGCAAGCAGC 1443
              | ||| |  ||||| |||| |||  | ||||  ||  | |||  ||| | ||||| |
Db   5474 ACGCCGGCGAAGACCATCGCGTGCGGTGCATCCGGAC-CCCAAGGTCACCGGC 5532

Qy   1444 AGCAGTGCCTCACGCCCGCCATCGTTACACCGGCGGTCCTCCGGACGTGGCGCGC 1503
              | ||||| | | ||||   | || ||  |||| | ||| |  | ||| || ||
Db   5533 CTGGGCCACGCCCACGTCGTCCCGGACTCGGCGCGCGCCCCGACGCCCCGGCCCC 5592

Qy   1504 CGCGGCGGGTGCGGGTGAACTGGTGGTCGTTGAGGGAGACCTTCGGAGCCCC 1563
              ||||||| ||||| ||||||  |||  |||  | | | ||  |  | |
Db   5593 CGCGGCGGGTGCGGGTGGTGGGGGGGAACCTGACC 5652

Qy   1564 CTGGTACCCTAACCGCCAGCACATTGACGAGCTCTTCTTCACCAAGACNGCACC 1623
              | | |||  |  ||| || ||  |||  | | | || ||| ||| ||||
Db   5653 ATCCCCGGGCGGGGAACCTCAGC---- 5686

Qy   1624 TGGTGGACTACTACCACCCCAAGGCGCGCAGAGACACGTGTACGGGCCGGGCACGT 1683
              |||  ||  || ||||||| || |  || ||  ||||| ||||||| ||| |||
Db   5687 AGGGCCACCAGCTCGAGCGCCAAGGCGCGGTGTACGGGCTGCGACGGTGC 5746

Qy   1684 ACACCGCCGACGAGGAGTACCGACTTCGGCTTCGGGCGGAGACTCGGGCAAAGG 1743
              || ||| |  || ||| |||| || ||| | | |  |  ||||||  |||||
Db   5747 CCGCCGCAGGGCGAGGTACCGCCGCCATGCGGCGTCCCACCGGGTGCAG- 5805

Qy   1744 CGTACGTAGACGACCAGCTCGTCGTCGACAACGCCACCAAGCAGGTCCCGGCGGATGCGCT 1803
              ||   | |||  |||| || |||| | ||  | | ||||| |   |  |
Db   5805 ----------------CTCGGCAGCCACA 5818

Qy   1804 TCTTCGGCTCCGCCACCGGGAGAGAGGGCCATAAATCTCGTCAAGGGCAACACGT 1863
              | || |||  |  | | |  | ||  ||| ||  || ||| |  || || | ||
Db   5819 CCATCGAGCGCGGTCAGGTCTACGCAAGGTGACCAGCCCGTCCTCAAGCTGACCAAGG 5878

Qy   1864 ACAAGTTCAAGATCGAGTTCGGCGTCCGCCACCACTACACCTCCAAGGCGGCGACACCATCG 1923
              | || ||   |||| ||| |||    ||  ||  ||  | ||  || || || ||  ||
Db   5879 GCACGCACAAGCT----------------CACGATCTCGGGCTTCGCGGTGA 5914

Qy   1924 TCCCCGGCCACGGCTCCCTCTCGCCGGTTCGGGGGGGCTGCCAAGGTCATTGACGACCAGGGCCGAAA 1983
              || ||| || || || | || ||| || ||||||||  | | || |||  ||||  ||||||
Db   5915 GTGCGCACCCCGGCTCTCCCTGGAGCTGGCGGTACGGCCGGACGGCGAGA 5974

Qy   1984 TCGAAAGTCCGTGCCTCGGCCTGCCCTCGCAAGGGACCACAGCCACAGACACCTCATCATCTGCGCGGGCGGCCTTA 2043
              |||| ||   |||  || |||  |  ||   |||  ||  ||| || |  || ||
Db   5975 TCGGCGAAGGCCGTGAGTCGGCGGAAGCGCCGTACGGGCGGGTCGTTTTGC------ 6026

Qy   2044 ACGCCGACTGGGAGAGACTGGGGGCGCGCCCGGAGCATGAACGTCCCCGGCGTGCTGG 2103
              | | | |    | |  ||  ||||||||| || || |||  ||||  |  |||||
Db   6027 -CTACGACGACGCCGAGGGCGTCGAGAACGGCGCCGAACCGTCGTGTCGCCGGGGTACGCAGG 6085

Qy   2104 ACCAGGCTCATTGCCCGACGTGTGGGCCGCGGCGGAACCAAACACCGTCTGCTGTCTCAGGACGG 2163
              | ||||| | |  ||| |  |||   ||||||| ||  | | |  |||
Db   6086 ACAAGCTGATCTCGGCCTGTGTCGCTGGAGCGCCAACCGGATCGTGGTCCTCAACACCG 6145

Qy   2164 GCACCCCGAGGAGAGATGCCCTCGGGCTTCGGACTTCATTGGGCTGCGCATCAGGCCGTGGTACG 2223
              |||| |  |   |||  |||  |  |||| | || |||||   |||  ||| |||
Db   6146 GTTCGGTGCGGGTGCTGATGCCCGGGACTACTGCTGTCCAAGAGAGCCTCGACCCCGTA 6205

Qy   2224 GCGGCAACGAGCACGGGCAACTCCAGTCGTCGTTTGGGACTACAACCCCTCG 2283
              ||  | || |  | |||  ||  ||| |||| ||||  ||||   ||||||
Db   6206 CGGGCGCCGAGGCCGGCCACCGCTCTCTACGGTGACGTCAACCGCTCGAGCG 6265

Qy   2284 GCAAGCTGTCCCTCAGCTTCCCC------------AAGGCCTTGCAGGACAACCCCGGTTTC 2334
              ||  ||| |  | |  ||||
Db   6266 GCAAGCTCAGCGACGAGCTTCCGCGGGCGCCGAGGAACACAGCGTCGCCGGGGGACCCGA 6325

Qy   2335 TCAACTTCCGCACCGAGGGCCGGCCGGCGGCGCTGTGTAGGCGGAGGACGCTCTAGGTCGGGGGTACA 2394
              ||  ||| |  | ||||   |  | ||||| | ||  |||  |   | |||  ||| | |
Db   6326 CAAGCTACCCGGCGGCGTGCCGGCCGACGCAACCCCAGCAGACACCACGCTCACCGTCGCGGGTACC 6385
```

```
Qy   2395 GGTACTACGAGTTTTGCGACAAGGAGCTCAATTTCCCCTTTGGCCACGATGCCATCCACGTCGGGTACC 2454
              |||   |  ||| ||  ||| | |  || | ||| |||||  |  |||||||||||  | |||
Db   6386 GCTCGGTTCGACAAGGAGAAGAACGTCAAGCCTGTTCCCGGTACTCCACGCCTCGGTCGCCTCTACA 6445

Qy   2455 CCACTTTTGCCTTTTCCAAATCTCTCCGTGTTCTCACAAGGAC---GGCAAGTCGAGGGTGT 2511
              || |||  || ||| ||  |  | | || |  |   | |||    |||||  | | |||
Db   6446 CCTCGTTCACGCAGAGCGCCCCGACCGTCGTGCGTACTGCGTACGGGCCACGGGTGGTCTTCACGTCA 6505

Qy   2512 CCCTCTCCGTGAAGAACCAACCGGTCCGGTCCCGGCCACAGGTGGCCCCAGTCTTCACGTCA 2571
              ||   | ||||| |   |||  | | |  | ||  |||||| | |||  |||| | ||||
Db   6506 CGGTCACGGTCCGCAACAGCGGAAGCGCGCGCGGCGCAGGAGGTCGTCTCCAGGGTACCTCG 6565

Qy   2572 AGCCCCTCCAACGCGCCCAAGAGATTAACCGGCGCCCGTCAAGGACGTCAAGGCGAAAGG 2631
              || |||| || |  |||   |  || |||||||  || | |||  | | |||  | ||
Db   6566 CGGTCACGGTCCGCAACGGGCTCCGGCAGGAGGTCGTCTCCAGGGTACACGAAGG 6625

Qy   2632 TCGAACTGCAGCCCGGCGAGGACGAAGGCGGGTGAC 2665
              || ||| ||||||  |||  || || |||||||||
Db   6626 TCTCGCGTCGCGCGGCGGAGGCGGAAGAGACGGTGAC 6659
```

```
FT   .
FT   CDS          /tag= h
FT               /product= "PikB gene cluster protein #8 (AAY77209)"
FT   CDS          complement (12342..13799)
FT               /tag= i
FT               /product= "PikB gene cluster protein #9 (AAY77210)"
FT   CDS          complement (13706..15043)
FT               /tag= j
FT               /product= "PikB gene cluster protein #10 (AAY77211)"
FT   CDS          15404..15574
FT               /tag= k
FT               /product= "PikB gene cluster protein #11 (AAY77212)"
FT

PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;     99WO-US14398.
XX
PR   26-JUN-1998;     98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77204, AAY77205, AAY77206, AAY77207, AAY77207, AAY77208,
DR   AAY77209, AAY77210, AAY77211, AAY77212, AAY80998, AAY80999.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Disclosure; Figure 32; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the production
CC   of biologically active macrolides. The macrolide biosynthetic proteins
CC   are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC   narbomycin. The alternative termination of polyketide synthesis may be
CC   useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC   monomers. The compounds produced by the recombinant host cells are useful
CC   as biopolymers, e.g., in packaging or biomedical applications, to
CC   engineer PHA monomer synthases or to prepare biologically active agents,
CC   such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC   chronic obstructive pulmonary disease as well as other diseases involving
CC   respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC   antibiotics which are active against a variety of organisms, e.g.,
CC   bacteria, including multi-drug resistant pneumococci and other
CC   respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC   protection agents (e.g., fungicides or insecticides) via expression of
CC   polyketides in plants. The present sequence represents the desosamine
CC   biosynthetic gene cluster (pik) from Streptomyces venezuelae ATCC
CC   15439, as given in figure 32.
XX
SQ   Sequence 13613 BP; 1858 A; 4733 C; 5092 G; 1930 T; 0 other;
```

```
Query Match             5.6%;  Score 167.2;  DB 21;  Length 13613;
Best Local Similarity  50.4%;  Pred. No: 4.2e-18;
Matches   496;  Conservative    0;  Mismatches  473;  Indels  15;  Gaps   3;

Qy   1697 GGACTGCACCTACGAGCTCGGCCCTCGTCGTTCGCGACGGCACGCAAAGGCGTACGTAGACGA 1756
           || |||| ||| |  |  ||||| |  | ||    ||| |||  |  |||| ||  |  ||
Db   3516 GGCGTTCAACCAGGCCCACCAGCTCGAGCGCGGGCCGGCGGGGGGGCGGCCCGCGGGGCAC 3575

Qy   1757 CCAGCTCGTGTCGACAAGCGCCACCAAGGCGAGGTCCCGCGCGGATCGCCTTCTTCGGCTCCGC 1816
           || |||| |||  | || | | ||| |||||||  |||| | || | | || |||  || |||
Db   3516 GGCGTTCAACCAGGCCCACCAGCTCGAGCGCGGGCCGGCGGGGGGGCGGCCCGCGGGGCAC 3575
```

```
Db   3576 GCTGACCGTGCCGCCGACGGCGAGTACCGGCGATCGCGGGTGCCACCGGTTACGG 3635
Qy   1817 CACCCGCGGGAGGAGACGGGCGCGGCCGCGATCAATCTCGTCAGACAGGTCCAAGAT 1876
            ||| |  ||| || || || || |||  | ||| | | | | |  || | | ||| ||
Db   3636 CACGGTGCAGCGTCGGCAGCAGCCACCCATCGAGCGGCGTCAGGTCATCTACGGCAAGAGTGAGCAG 3695

Qy   1877 CGAGTTCGGCGTCCGCACCTACACCCTCAAAGGCGACAACCATCGTCCCGCGCCACGG 1936
            | |  ||| || |  || |||| ||| || ||| || || ||| | ||| || |||| |||
Db   3696 CCCGGCTCCTCAAGCTGACCAAAGGGGACGACGCACAAGCTCACGATCTCGGCTTCGGCGATGAG 3755

Qy   1937 CTCCCTCCGCGGTCGGCCGCTGGCTCAAGGTCATTGACGAGCCGAAATCAAAAAGTCCGT 1996
           || | || || |  |||| | || ||||| |||| | |   | | |  |||| | |||
Db   3756 TGCCACCCCGGTCTCCCGTGGAGCTGGGCTGGGTGACGCCGGGCGGCGACGGCGACGAT 3815

Qy   1997 CGCGCTTCGCCAAGGGAGCGGCACGACCAGGTCATCATCTCGCGGGGCCTTAACGC---CGACTG 2053
           ||||| |||  ||| ||| | | || |||| |||||||  || |||||| | || |   || ||
Db   3816 CGCGAAGGCCCTTGGACTGGGCGCCAACCCGGAACAGGTCCTCAACACCGGTTCGTCGGT 3875

Qy   2054 GGAGACCGAGGGCGTCGGGGGGGGGGGGCACCCGGCGAGCAATGAGAAGGTCCCCCGGGCTGCATCGGACGAGCTCAT 2113
            || ||||||||| |   |||  || |||| || ||  ||  | | ||| || || ||| | |  ||| || |||
Db   3876 CGGCACCGAGGGCGTCGACACGTGCCGAAACCTGTCGACAAGCTTCGTCACAGGC 3935

Qy   2114 TGCCGACGTGGCGCCGGCGAACCCAAAACAACGTCGTCGTGCAGGGCGCACCCCCGA 2173
           ||    || || || ||| | |   ||||| | |  || || | |||| | || | ||
Db   3936 CTCGGCGTGTCCGCGGACGGCACAACCCGAACACGGATCTGGGTGTCCTCAACACCGGTTCGTCGGT 3995

Qy   2174 GGAGAATGACCCTCTGGCGTCGAGACGCCACCAGCCCCGCCCGTCGTCATCCAGGCCTCAGGGTGCATAGGTTACGACCTACG 2233
            ||||  || |  | ||| |||| ||| || |||| ||||| |  |||| |||| ||| | ||| |||  ||||| |
Db   3996 GCTGATGCCGTGGCCTGTCCAAGACCCGCGCGCGGTCCTTCGCGGACATGTCGGTACCCGGCGCCAGGC 4055

Qy   2234 GACGGGCCAACTCCATTGCCGACGTCGTTCTTTGGGGACTACAAACCCTCGGGCAAGCCTGTC 2293
           ||||||| ||| || |||| || | |  | |  ||||  ||| |||||| ||||| || |||
Db   4056 GGGCCGCCGAGGCCACCGCCGCCGCTGCTCTACACGGTCGACGTCAACCCGAGGGGCAAGCTCAC 4115

Qy   2294 CCTCAGCTTCCCC-------AAGCGCCTGCAGAGACAACACCCCGGCGTTTCTCAACTTCCG 2344
           || |||| || ||       ||||| ||| || || |||| || ||| ||| |  |  |||
Db   4116 GCAGAGCTTCCCGGCCCGGAAGACACAGCCCCGAGAGAACGCGGCGCCGGAGCCCGAGGCGTACC 4175

Qy   2345 CACCGAGGGCCGGGGGCGTCTACGGCGACGGAGGAGGGTCGTACAGGGGTACAGGGGTACACGA 2404
           |  || |||||||| |||||| ||||| || || ||| ||||| | |||| |||| |||| ||
Db   4176 GGGCGTCGACAAACCAGCAGACGAGTACCGGCGAGGGCATCCACGGTCGGGTACCGCTGGTTCGA 4235

Qy   2405 GTTTGCCGACAAGGACGTCAATTCCCCTTTGGCCACGGCCGCTGTCCTACACCACTTTTGC 2464
           | ||| || || ||  || |||| ||| |||| ||| |||| ||  ||||  ||| ||||
Db   4236 CAAGGAGAACGTCAAGCCCGTCTGTCCCGGTCCCGCACCCGCTGGTTCAC 4295

Qy   2465 CTTTTCCAATCTCTCCGTGTCTCACAAGGAC---GGCAAGCTGACCGTGTCCCTCTCCGT 2521
           || | |  | | | |  ||||| ||| |||   ||| |||  ||| || ||| || ||
Db   4296 GCAGAGCCGCCCGAGGTCGTGCGGTACGTCCACGGTCTGCAGAGGTCACGGTCAGCGGT 4355

Qy   2522 GAAGAACACCGGCTCCGTGCCCGGCGCACAGGTGGCCAGCTCTACGGTCAAGCCCCTCCA 2581
           ||||  || ||| || || ||||  ||| || || |||| | || ||| ||| || || |
Db   4356 CCGCAACACGGGAAGCGGCGGGCCGGAGGTCGTCGGTGCCGAGGTCGTCGGTGCCGAGCC 4415

Qy   2582 AGCCGGCCAAGATTAACCGCCCCGTCAAGGAGCTTCGCAAAGGTCGAAACTCCA 2641
           || || |  || ||| ||| ||| | ||||| |  ||| || |  ||  ||| |
Db   4416 GAACGTCGACGGCTCCGCAGCAGAAGAGCTCGTGGGCTACACGAAGGTCTCGCACGGT 4475

Qy   2642 GCCCGGCGGAGGACGAAGGCGGTGAC 2665
           ||| || || ||  ||| ||||||
Db   4476 CGCGGGCGGAGGCGAAGACGGTGAC 4499
```

```
RESULT 11
AAT32999
ID   AAT32999 standard; DNA; 2256 BP.
XX
AC   AAT32999;
XX
DT   18-OCT-1996  (first entry)
XX
DE   Chimeric thermostable beta-glucosidase coding sequence.
XX
```

```
KW   Chimaera; beta-glucosidase; Cellvibrio gilvus; Agrobacterium tumefaciens;
KW   homology; decomposed cellulose; glucose; cellobiose; cellotriose;
KW   cellotetraose; cellopentaose; ds.
XX
OS   Synthetic.
XX
FH   Key            Location/Qualifiers
FH
FT   misc_feature   1..2076
FT                  /*tag= a
FT                  /note= "sequence derived from Cellvibrio gilvus
FT                  beta-glucosidase gene"
FT   misc_feature   2077..2256
FT                  /*tag= b
FT                  /note= "sequence derived from Agrobacterium tumefaciens
FT                  beta-glucosidase gene, encodes amino acids
FT                  759-818 of A.tumefaciens beta-glucosidase"
XX
PN   JP08131168-A.
XX
PD   28-MAY-1996.
XX
PF   09-NOV-1994;   94JP-0299049.
XX
PR   09-NOV-1994;   94JP-0299049.
XX
PA   (NORQ ) NORINSUISANSHO SHOKUHIN SOGO.
XX
DR   WPI; 1996-311636/32.
DR   P-PSDB; AAR97199.
XX
PT   Thermostable chimaeric beta-glucosidase - useful for prodn. of
PT   cellobiose from decomposed cellulose
XX
PS   Claim 2; Page 5-8; 9pp; Japanese.
XX
CC   This is the nucleotide sequence encoding a chimaeric beta-glucosidase
CC   enzyme composed of amino acids 1-692 of Cellvibrio gilvus
CC   beta-glucosidase and amino acids 759-818 from the Agrobacterium
CC   tumefaciens beta-glucosidase, replacing amino acids 693-752 of the
CC   Cellvibrio sequence. The chimaera was constructed by isolating the
CC   C.gilvus beta-glucosidase gene and comparing its sequence with the
CC   corresp. genes from several other species, e.g. Hansenula anomala,
CC   Ruminococcus albus, Butyrivibrio fibrisovens, Agrobacterium tumefaciens,
CC   etc. The homologous sequences from the other species could then be used
CC   to generate chimaeric glucosidase genes. The novel chimaeric protein
CC   acts on decomposed cellulose to generate glucose and cellobiose and has
CC   a reduced ability to decompose cellobiose as compared to cellotriose,
CC   cellotetraose or cellopentaose as substrate. The novel enzyme has an
CC   optimum pH of 6.0 and temp of 40-45 deg. C, is stable in pH 4-9 and at
CC   35 deg. C.
XX
SQ   Sequence 2256 BP; 370 A; 786 C; 763 G; 337 T; 0 other;
```

```
Query Match     52.4%;
Best Local Similarity   5.4%;
Matches  449;  Conservative 0;  Mismatches 375;  Indels 33;  Gaps 3;
Score 161;  DB 17;  Length 2256;  Pred. No. 3.8e-17;
```

```
Qy 1936 GCTCCCTCCGCGTTCCGGCGGCGCTCCAAGGTCATCTCTGCGCGGCGGGC 1995
        |||  |||  |||   ||| |||  | |||   |  ||
Db 1346 GCCGGATCCAGGCGCCCAGGCCCCGGAAATCGGAAAGTCGTTCGAACGGCGCGGGACCGAC 1405

Qy 1996 TCGCCCTCGGCCAAGGAGCACGACCAGGTCATCATCTGCGCGGCGGCCTTAACGCGACTGGG 2055
        |||     ||| |||     ||| | |||   | ||
Db 1406 CCCGCCGCGCGCTGGCCGTGGGGCCGCAGCTCGGGCCGCGAGACTCATTG 1465

Qy 2056 AGACCGAGGGGCGGCCGGCCGACGCCGCGGAGACTATGAGAAGCTCATTG 2115
        ||||| |    |||  ||  |||     |||  ||  ||
Db 1466 TCGGCGAGGCCAACAGCGCCCAGACGCTGCGCCGCGCCAGAAGAGCTGATCA 1525

Qy 2116 CCGACGTGGCCGCGCGGCGGACCCAAAACACGTGGTCGTCGTCATGCAGGACGGGCACCCCCGAGG 2175
        |||| |||  |||   ||||||   |  |  || ||| |  ||
Db 1526 CGTCGGGTGGCGGCGCGCGCAACGGCGCCCACCGTGTCGTTGTGTCTGCAGGACCGGGTCA 1585
```

```
Qy 2176 AGATGCCCTGGGCTGCGAGGCCCACGGCCCGCCGCCGTCATCCAGGCGTGGTACGGCGCGGCAACCGAGA 2235
        ||  ||||||||   |||  |||| |||  ||  ||    |||||||    ||||    ||||  |||| ||||
Db 1586 CCAATGCCCTGGCTGGCGGCGCGTTCGGCGGCGGTTCCGGCGGCTGCTGGGAAGCCTGGTATCCGGCCAGCG 1645

Qy 2236 CGGGCAACTCCATTGCCGACGTCGGTCGTTGGGGGACTACAACCCCTCGGCAAGCTGTCCC 2295
        || ||||  ||||    | ||||||  | ||| ||||  || |||  || |||||||||
Db 1646 GCGGCGAGGGCGATCGCGAAATGTGCTCGTTGTTGGGCGGTCAATCCTCGTCGGGCCGGAA 1705

Qy 2296 TCAGCTTCCCAAGCGCCTGCAGGACGAACACCCGCGGTTCTCAACTTCCGCACCGAGCCG 2355
        |||     ||  |||  | || | |||  ||||||   ||   ||  | || ||| ||
Db 1706 CCACCTTCCCGCAATCCGACAGCAACTGCCCGGGCCCCGAAACCTCGATGGCGATCCGGAA 1765

Qy 2356 GGGCGCACCGCTGTACGGCGGAGGACGGTCTAC---------------GTCGGGTACAGGGTACT 2400
        || || |||  || |||||  | |||||  |||               | ||| |||  ||| |
Db 1766 ACCCGAGCGTGCAGTTCGGCGTCGGCTGCGACTACCACGACGAAGGTTGCGGCGCGGGAGCTAGAAGTGGT 1825

Qy 2401 ACGAGTTTGCCGACAAGACGACGTCAATTTCCCCCTTTGGCCACCGCCGCGTGCCTACACCACTT 2460
        ||  ||||  ||||| | |||  || | |  ||| ||| || ||| | |||| || ||| | | |
Db 1826 TCGACCTGAAAGGCGCCACAAGCACCGCTGTTCCCGTTCGGCGACCGCCTGTCCTACAACAACT 1885

Qy 2461 TTGCCTTTCCAAATCTCTCCGTGTTCTCACAAGGACGACGAAGCTGAGCGTGTTCCCTCTCCG 2520
        || |||    | |  |  ||| | || ||| || ||| || | |||| ||| ||  | |||||
Db 1886 TCGGGTACTCCGGTGTCTGTCCGGCCAGCTCAAAGGAATGGCGGGGTTCAAGG 1945

Qy 2521 TGAAGAAACACCGGGTTCCGTGCGCCCGGGCACAGTGGCCCCAGCTCACTTCAAGGCCCCTCC 2580
        ||  |  |||  |||  ||| |||||||| ||   |||| ||||||| | ||||| | |  ||
Db 1946 TGACCAAACACCGGCAAACGGTGGGCCCGGCCCAGAGGTGCCCCAGGTGTACCGGCCCGGAT-- 2003

Qy 2581 AAGGCGCCAAGATTAACCGCGCCCGTCAAGGAGCTCGCAAAGGTCGGAACTGC 2640
        ||| || ||  |  ||| | ||| | || | ||||  | |||| |  |  ||
Db 2004 ----GTCCACCAAATGGGAGGCCGGGTGACAATCGAGGAGCGAGAGTACGTGCGCT------- 2059

Qy 2641 AGCCCGGCGGAGACCGAAGGCGGGTGACAAATCGAGGAGCCGAGAAGTGCCCGC 2693
        |||||||| ||||| ||| | |||  | |||   ||||| ||| |||||
Db 2060 TGCCGGCGGAGACCAAGCCGGGCGGCGGGCCGACCGGCTACGGCCCGGTGGGCG 2119

Qy 2694 ------GCGTATTTGATGAGGAGGCGGATCAGTCGGTGCTGGTGGTTGACTATGAGG 2748
              ||| ||| |||  | ||| || ||| | |||  | |||   | ||||||
Db 2120 ACTTGGCTTACTTCGATGTGCGAGGCGGCGGCGTCGGTTCCGGCGGCGGCAAGTACGAGC 2179

Qy 2749 TTATCGTGAGGGACAGC 2765
        | ||||||  |||||||
Db 2180 TGATCGTGGCGGCCAGC 2196
```

```
RESULT 12
AAA09469/c
ID   AAA09469 standard; DNA; 50937 BP.
XX
AC   AAA09469;
XX
DT   29-AUG-2000  (first entry)
XX
DE   Streptococcus oleandomycin gene cluster.
XX
OS   Streptococcus antibioticus.
XX
KW   oleandomycin; oleandolide; polyketide synthase; oleA; oleAI; oleAII; oleAIII;
KW   PKS; type I; 8,8a-deoxyoleandolide synthase; modular; ketosynthase;
KW   acyl-transferase; acyl carrier protein; inactivated; polyketide;
KW   macrolactone; antibiotic; motilide; erythromycin; ss.
XX
FH   Key            Location/Qualifiers
FH
FT   CDS            152..1426
FT                  /*tag= a
FT                  /label= oleI
FT   CDS            complement (1528..2637)
FT                  /*tag= b
FT                  /label= olen2
FT   CDS            complement (2658..4967)
FT                  /*tag= c
FT                  /label= oleR
FT   CDS            5772..18224
FT                  /*tag= ?
FT                  /label= oleAI
```

```
FT
FT
FT   CDS             /*label= ORF1
FT                   /product= 8,8a-deoxyoleandolide_synthase_1
FT                   /*tag= e
FT                   18267..29717
FT   CDS             /*label= ORF2
FT                   /product= 8,8a-deoxyoleandolide_synthase_2
FT                   /*tag= f
FT                   29787..40346
FT   CDS             /*label= ORF3
FT                   /product= 8,8a-deoxyoleandolide_synthase_3
FT                   /*tag= g
FT                   40625..41830
FT   CDS             /*label= olePl
FT                   /*tag= h
FT                   41878..43158
FT   CDS             /*label= oleGl
FT                   /*tag= i
FT                   43163..44443
FT   CDS             /*label= oleG2
FT                   /*tag= j
FT                   44433..45173
FT   CDS             /*label= oleM1
FT                   /*tag= k
FT                   45251..46411
FT   CDS             /*label= oleY
FT                   /*tag= l
FT                   46491..47714
FT   CDS             /*label= oleP
FT                   /*tag= m
FT                   complement (47808..49517)
FT   CDS             /*label= oleB
XX
XX
XX
PN   WO20026349-A2.
XX
PD   11-MAY-2000.
XX
PF   22-OCT-1999;        99WO-US24478.
XX
PR   29-OCT-1998;        98US-0106100.
PR   16-FEB-1999;        99US-0120254.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Betlach MC,  Shah SK,  McDaniel R,  Tang L;
XX
DR   WPI; 2000-365602/31.
DR   P-PSDB; AAY92707, AAY92708, AAY92709.
XX
PT   Recombinant DNA compound encoding oleandolide polyketide synthase for
PT   synthesizing polyketides comprising a coding sequence for a domain of a
PT   loading module or any one of extender modules
XX
PS   Disclosure; Page 14-26; 86pp; English.
XX
CC   This is part of the Streptococcus antibioticus oleandomycin gene cluster.
CC   The oleandolide polyketide synthase (PKS), also known as
CC   8,8a-deoxyoleandolide synthase, is encoded by three open reading frames
CC   (ORF), designated oleAI, oleAII and oleAIII. The PKS is a type I
CC   "modular" enzyme, where each ORF encodes 2 extender modules and
CC   the first ORF also encodes the loading module. Each module is composed
CC   of at least a ketosynthase (KS), acyl-transferase (AT) and an
CC   acyl carrier protein (ACP) domain. The oleandolide PKS loading module
CC   contains an inactivated KS, called KS-Q, where Q is the abbreviation for
CC   glutamine, present instead of the active site cysteine required for
CC   activity. The large multifunctional PKS enzymes catalyze the biosynthesis
CC   of polyketide macrolactones through multistep pathways involving
CC   decarboxylative condensations between acylthioesters followed by cycles
CC   of varying beta-carbon processing activities. The macrolide product of
CC   the PKS, 8,8a-deoxyoleandolide, is further modified by epoxidation and
CC   glycosylation to yield oleandomycin, an antibacterial polyketide. The
CC   invention concerns an isolated recombinant DNA compound, comprising a
CC   coding sequence for a domain of loading module or any one of extender
```

```
CC   modules 1-4 or 1-6, including an oleandolide PKS operably linked to a
CC   promoter. Also discussed are recombinant oleandolide PKS in which the
CC   module 1 KS domain is inactivated by deletion or other mutation. In
CC   particular, the inactivation is mediated by a change in the KS domain
CC   that renders it incapable of binding substrate (the KS1-o mutation),
CC   rendered by mutation in the codon for the active site cysteine. The
CC   oleandolide PKS is useful for synthesizing polyketides, which are useful
CC   as antibiotics and motilides. Heterologous expression of oleandolide PKS
CC   in host cells such as Streptomyces coelicolor and S. lividans is also
CC   made possible. Unmodified oleandolide compounds can be provided to
CC   cultures of Saccharopolyspora erythraea and converted to the
CC   corresponding derivatives of erythromycins A-D.
XX
SQ   Sequence 50937 BP; 6672 A; 16253 C; 19272 G; 8740 T; 0 other;
```

```
Query Match             4.7%;  Score 138.6; DB 21;  Length 50937;
Best Local Similarity   54.0%;  Pred. No. 1.8e-13;
Matches  354;  Conservative   0;  Mismatches 289;  Indels  12;  Gaps   3;

Qy   2048 CGACTGGGAGACCGAGGGGCCGACCGAGCATGAAGCTCCCGGCGGCGTTCTGGACCA 2107
           ||||   | ||||  || |||||||| ||||| | |||||||||| |||| | |||||
Db   3417 CGACGACGGGTCGGAGGACGGGACGGAGCCGCACGTCGCTGTCCCTGCCCGGCGAGGA 3358

Qy   2108 GCTCATTGCCGACGTGGCCCGGCGCCGGCGAAACCCAAACACCGTCGTGCATGCACGACGGGGCCAC 2167
           || |||  |||  ||| ||| || ||||||    || ||  |  ||| ||| |||  |||  |||||
Db   3357 CCTGATCAGCCGCCGGTGGCCCGCCGGGTCAACCCACCAGAACCGTGGTGGTGTGCTCAACACGGGCTC 3298

Qy   2168 CCCCGGAGGAGATGCCCTGGCTCGAGCACCGCCCGCGGTCATCCGAGGCCTGGTACGGCGG 2227
           ||||||||||||||||||||||||||||||||||| | || ||||| | ||||||||||||
Db   3297 CTCCGTCACCATGCCCCTGCGTTGCGTGAGGAAGAACCGCCGCCGGGTTGCTCACCATGTGGTTACCGGG 3238

Qy   2228 CAACGACGCGGGCAACTCCATTGCCGACGTCGTCTTTGCGGACTTACAACCCTCGGGCAA 2287
           || ||||  ||||||| ||  | ||| ||  ||||| ||  |||||||||||||||||||
Db   3237 CCAGGCGCGGGCGGAGGGGACCACCGCCCGACGTGCTCGTTCGCGGAGCGCACCACCAGGAGGCCG 3178

Qy   2288 GCTGTCCCTCAGCTTCCC-----CAAGCGCCTGCAGGACACAACCCCGCGTTTCTCAACTT 2341
           |||||  |||| | | ||     || |||| ||||||| |||  || ||| ||  ||||||
Db   3177 GCTGACCCGACACCTTCCCGGCGCCGAGGAGCAGGACGCGGTTCCTTCGGCCACAGGGGCCACGGG 3118

Qy   2342 CCGCACCGAGGCCCGGCGCCACGCGCTG---TACGGCGGAGGACGGTCTACGTCG3GTACAGGTA 2398
           || ||| || || ||||| | || ||    ||| || |||| ||||| ||| ||   ||  |||
Db   3357 CCTGATCAGCCGCCGGTGGCCCGCCGGGTCAACCCACCAGAACCGTGGTGGTGTGCTCAACACGGGCTC 3298

Qy   2168 CCCCGGAGGAGATGCCCTGGCTCGAGCACCGCCCGCGGTCATCCGAGGCCTGGTACGGCGG 2227
```

```
Qy   2399 CTACGAGTTGCCGACAAGGACGTCAATTTCCCGTTTGGCCACGGCCTGTCCTTACCACCAC 2458
           | ||||| |||||| |||||||||||||||||||||||||||||||||||||| ||||||
Db   3058 CACCGCGTGCGCAACAACCGGTACGCGCACCGGCAAGGAGGTCCGCAGGTCTTACGTGGGGGCC 2998

Qy   2459 TTTTGCCCTTTTCCAATCTCTCCCGTGTCTCACAA---GGACGGGCCAAGCTGACCGTGTCCCT 2515
           |||||||  | ||  | |   |||||||||  |  |   |||  ||||||   ||| |||||||
Db   2997 CTTCGACTACCGGGACCTGAAGGTGACCGGCGAGGCGGTCAAGGGCCTTCGACGTCTGCCTT 2938

Qy   2516 CTCCGTGAAGAAACACCGGCGTCCGGCGGCCACAGGTGGCCCCAGCTCTACGTCAAGCC 2575
           | |||| | ||||||||||| |||||||||||| |||||||||||||||||||||||||
Db   2937 CACCGCGTGCGCAACAACCGGTACGCGCACCGGCAAGGAGGTCCGCAGGTCTTACGTGGGGGCC 2878

Qy   2576 CCTCCAAGCGGCCAAGATTAACCGCCCGTCAAGGAGCTCAAGGGCTTCGGCAAAGGTCGA 2635
           | ||| ||| ||||||||  ||||  |||||||| ||||| ||||||| || |||||||
Db   2877 GTCCCCGCACGTCGGGTTGGCGCGAGCCAAGCGTGCCCGGGCCTAAGGAAGGTCGA 2818

Qy   2636 ACTGCAGCCCGGCGGACGAAGGCGGTGACAATCGAGGAGCAGGAGGAGAAAGTACGTC 2690
           |||| || | |||| | |||| | || || | ||| |||||| ||| || |||| |||
Db   2817 GCTGCGCGGCCGGCGGCGGAGCCGGAGGACGGCTTGACCGTGGGCCAGGCCGGGCCCTC 2763
```

```
XX
KW   Glycosidase; thermostable; textile; food processing; pharmaceutical;
KW   detergent; baking; industry; Thermococcus; Staphylothermus;
KW   Pyrococcus; glucose; soluble oligosaccharide; ss.
XX
OS   Thermotoga maritima.
XX
FH   Key          Location/Qualifiers
FT   CDS          1..2166
FT                /*tag= a
FT                /product= Glycosidase
XX
PN   WO9725417-A1.
XX
PD   17-JUL-1997.
XX
PF   10-JAN-1997;       97WO-US00092.
XX
PR   13-SEP-1996;       96US-0712612.
PR   11-JAN-1996;       96US-0583787.
XX
PA   (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX
PI   Bylina EJ, Lam DE, Mathur EJ, Swanson RV;
XX
DR   WPI; 1997-372858/34.
DR   P-PSDB; AAW34558.
XX
PS   Claim 4; Fig 5; 82pp; English.
XX
CC   The present sequence encodes glycosidase isolated from Thermotoga
CC   maritima. The enzyme or its encoding nucleic acid sequence is used for
CC   generating glucose from soluble oligosaccharides. The enzyme can be
CC   used in the food processing, pharmaceutical, textile, detergent and
CC   baking industries. The enzyme is also used to treat lactose intolerance,
CC   as a diagnostic reporter molecule, in corn wet milling or in the fruit
CC   juice industry. The enzymes can be used to hydrolyse guar gum to remove
CC   non-reducing terminal mannose residues. The nucleic acids encoding the
CC   enzyme may be used to generate probes to identify similar sequences.
XX
SQ   Sequence 2166 BP; 668 A; 454 C; 578 G; 466 T; 0 other;
```

```
Query Match    4.6%;   Score 135.8;   DB 18;   Length 2166;
Best Local Similarity   53.5%;  Pred. No. 4.3e-13;
Matches  284;  Conservative   0;  Mismatches 247;  Indels   0;  Gaps   0;

Qy   545 GCTCGGTTCCACATTCAACCAAACTCTTCGCTCCGAAGAGGCAGGTAAGATGAGTGAGGGCAAAGA 604
Db   258 GCTCGGCTTCTACCTGGAACAGAGCACCTTCTGGAAGAAGAGTGGGAAAAGCCATGGGAGAAGA 317

Qy  .605 GGCCATCGCTAAGAGTGCGCARGTGATCCTCGCGCCCGACTAATCAACAATGCAACGCTCCCC 664
Db   318 AGTTAGGGAAATACGGTGTCGATGTGCTTCTTCGCACCTCGGATGAACATTCACAGAAACC 377

Qy   665 TCTCGGTGGACGGTCGAGTCGATTGGTGAGGATCCGTTCCTGCGGGCCTTGGGAGC 724
Db   378 TCTTTGTGGAAGGAAATTCGAGTACTACTCAGAAGAATCCTGTTCCGGTGAAATGGC 437

Qy   725 TGGCGCTCTCATCACCCGGCGGCATTCGAGCACTGGAGTCGAGTGGCGTACGATCAACGCACTTTTT 784
Db   438 TTCAGCCTTTGTCAAGGGAGTTCAATCTCAAGGGBTGGGAGCCTGCATAAAAACACTTTTGT 497

Qy   785 GTGCAATGATCAGGAGGAGACAGGGCGGCCATCATATGTTGTCAGAGACCATCGCGGAGGGGCCTCT 844
Db   498 CGCGAACAACCAGGAAAACGAAGGATAGGTTAGTGAGGACAGACGATACGTGTCCGAGCGGAGCCCT 557

Qy   845 CCGTGAAAATCTACCGCACTCCCGTTCCCAGATTGTGTGGCGAGACCTCCCACCCGGGGTGCGTT 904
Db   558 CAGAGAAAATATATCTGAAAGGTTTTGAAAATTGCTCTGCTCAAGAAAAGCAAGACCCTGGACCGT 617

Qy   905 CATGACGGGCGGTACAATGGCCATCAATGGCGTGTTGTCGGTGCAGCGAGGAGAACCCTAAATATCTTGA 964
Db   618 GATGAGCGCTTACAACAACTGAATGGAAAAATACTGTTCACAGAAGCGAATGGCTTTTGAA 677

Qy   965 TGGGATGCTTCTGAAAGGAAATGGCGGGTTGGGATGGCGTTAATCATGAGACGGACTCGGTTCGTACGGCAC 1024
Db   678 GAAGGTTCTTCAGGGAGAAGAAATGGGGAGGGGATTTGGCGGTTTCGGTGATGAGGCGACTGGGTTACGCGGG 737

Qy  1025 ATACAGTACCACAGAGAAGGCCGTTGTGCGAGGCCTTGCGACCTCGAGAGTGCCCCGG 1075
Db   738 AGACAACCCTGTAGAACAGCTCAAGGCCCGGAAACGATATGATCAATGCCTGG 788
```

```
RESULT 14
AAV36911
ID   AAV36911 standard; DNA; 2166 BP.
XX
AC   AAV36911;
XX
DT   21-DEC-1998  (first entry)
XX
DE   Thermotoga maritima MSB8-6G glycosidase gene coding region.
XX
KW   Glycosidase; MSB8-6G; thermostable enzyme; oligosaccharide;
KW   glucose; sugar; baking; textile; detergent; beta-galactosidase; ss.
XX
OS   Thermotoga maritima strain MSB8-6G.
XX
PN   WO9824799-A1.
XX
PD   11-JUN-1998.
XX
PF   08-DEC-1997;       97WO-US22623.
XX
PR   10-OCT-1997;       97US-0949026.
PR   06-DEC-1996;       96US-0056916.
XX
PA   (DIVE-) DIVERSA CORP.
XX
PI   Bylina EJ, Lam DE, Mathur EJ, Swanson RV;
XX
DR   WPI; 1998-362407/31.
DR   P-PSDB; AAW49862.
XX
PS   Claim 1; Fig 5a-b; 92pp; English.
XX
CC   This isolated polynucleotide comprises a coding region for
CC   glycosidase MSB6-6G (see AAW49862) from a Thermotoga maritima MSB8
CC   clone (6G) that grows optimally at 85 degc in high salt medium.
CC   The sequence shows 538 nucleic acid identity to beta-galactosidase
CC   B of Clostridium thermocellum. The invention provides 18
CC   polynucleotides (see AAV36907-24) coding for thermostable glycosidases
CC   (see AAW49858-75) having glucosidase, alpha-galactosidase,
CC   beta-galactosidase, beta-mannosidase, endoglucanase
CC   or pullulanase activity. Vectors and host cells are also claimed.
CC   A method is provided for producing the enzymes by recombinant
CC   techniques. A claimed method for generating glucose from soluble
CC   oligosaccharides comprises contacting a sample (selected from
CC   dairy products, fruit juice, detergent, textile, guar gum, animal
CC   feed, plant biomass or waste product) containing oligosaccharides
CC   (selected from maltose, cellobiose, lactose, sucrose, raffinose,
CC   stachyose, verbascose, cellulose, starch, amylose, glycogen,
CC   disaccharides, polysaccharides and pullulan) with one of the
CC   claimed glycosidases such that glucose is produced.
XX
SQ   Sequence 2166 BP; 675 A; 444 C; 581 G; 466 T; 0 other;
```

```
PT   Glycosidase enzymes from organisms of the genera Staphlothermus,
PT   Pyrococcus and Thermococcus - for deriving sugar from
PT   oligosaccharides, useful in the e.g. food processing, textile or
PT   baking industries.
```

```
RESULT 15
AAI99683
ID   AAI99683 standard; DNA; 4403765 BP.
XX
AC   AAI99683;
XX
DT   15-JAN-2002  (first entry)
XX
DE   Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
KW   Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW   variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS   Mycobacterium tuberculosis.
XX
PN   US6294328-B1.
XX
PD   25-SEP-2001.
XX
PF   24-JUN-1998; 98US-0103840.
PR   24-JUN-1998; 98US-0103840.
XX
PA   (GENO-) INST GENOMIC RES.
PI   Fleischmann RD, White OR, Fraser CM, Venter JC;
XX
DR   WPI; 2001-647261/74.
XX
PT   Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT   determining the nucleotide sequence of the strain at positions in the
PT   genome corresponding to positions where M. tuberculosis strains CDC
```

```
PT   1551 and H37Rv differ -
XX
PS   Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.
XX
CC   The invention relates to evaluating strain variation within and between
CC   different populations of the tuberculosis bacterial pathogen,
CC   Mycobacterium tuberculosis or related Mycobacterium by determining the
CC   nucleotide sequence of the first strain at positions in the complete
CC   sequence of the genome that correspond to positions that differ in the
CC   nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
CC   H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC   M. tuberculosis and has valuable application in the fields of
CC   tuberculosis genetics, epidemiology, patient treatment and epidemic
CC   monitoring.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at seqdata.uspto.gov/sequence.html?docID=6294328B1.
XX
SQ   Sequence 4403765 BP; 1447799 C; 1441301 G; 757371 T; 189 other;
```

Query Match   Best Local Similarity 53.5%;   Score 135.8; DB 19; Length 2166;
Matches 284; Conservative 0; Mismatches 247; Pred. No. 4.3e-13; Indels 0; Gaps 0;

```
Qy   545 GCTCGGTTCCACATTCAACCAAAACTCTGCCTGCCTGCGAAGAGGCAGGTAAGGATGATGGGCAAAGA 604
Db   258 GCTCGCTTCTACCTGGACAGAGAACCTTCTGGAAGAAGTGGGAAAAGCCATGGGAGAGAAGA 317

Qy   605 GGGCCATGGCTAAGAGTGCGCATGTGATCGCCGGCCCGACTATACAACATGCAACGGCTCCCC 664
Db   318 AGTTAGGGAATACGGTGTCGATGTGCTTCTTGCACCTGCGATGAACATTCACAGAACC 377

Qy   665 TCTCGGTGGACGTGGCTTCGAGTTCGGATTCCGTTCCTGGCGGGGCTTGGGAGC 724
Db   378 TCTTTGTGGAAGGAATTTCGAGTACTACTCAGAGAATCCTTGCTCCTTTCCGGTGAAATGGC 437

Qy   725 TGGGGCTCTCATCCGCGGCATTCAGAGCACTGGAGTTGCACGGCTACGATCAAGCACTTTTT 784
Db   438 TTCAGCCTTTGTCAAGGGAGTTCAATCTCAAGGGGTGGGAGCCTGCCATAAAACACTTTGT 497

Qy   785 GTGCAATGATCAGGAGGAGCAGGCGCATGATGGTCAGAGCATCGTCACGGAGCGGGCTCT 844
Db   498 CCGGAACAACCAGGAAAAACGAACAGATGGTAGTGACACGCATCTGTGTCCAGCGCAGCCCT 557

Qy   845 CCGTGAAAATCTACGGCACTCCCGTTCCAGATTTGCTGTGCGAGACTTCCCAGCCGGGGTGCGTT 904
Db   558 CAGAGAAATATATCTGAAAGGTTTTGAAATTGCTGTCAAGAAAGCCAAGAACACCTTGGACCGT 617

Qy   905 CATGACGGCGGTACAAATGGCATCAATGGCGTGTCGTGCAGCGAGAACCCTTAAATATCTTGA 964
Db   618 GATGAGCGCGTTACAACAAACTGAATGAAATGGAAAAATACTGTTCACAGAACGAATGGCTTTTGAA 677

Qy   965 TGGGATGGCTTCGAAAGGAAATGGGGGTTTGGGGGATGGGCCTAATCATCATGAGCGGACTGGTTACGGCCAC 1024
Db   678 GAAGGTTCTCAGGGAAGGAGAAATGGGGGATTTGGCGGTTTCGTGATGAGCGACTGGTACGGCGGG 737

Qy  1025 ATACAGTACCACAGAGAAGCCGTTGTTGGGCAGGCCTCGACCTCGAGATGCCCGG 1075
Db   738 AGACAAACCCGTAGAACACGTCAAGGCCGGAAACGAATATGATCATGCCTGG 788
```

Query Match   Best Local Similarity 52.0%;   Score 111.6; DB 22; Length 4403765;
Matches 345; Conservative 0; Mismatches 289; Pred. No. 4.9e-09; Indels 30; Gaps 3;

```
Qy  2061 GAGGGCGCCGCCGACCGGCCGGAGCAATGAAGAGCTCCCCGGGCCGCGTGGTGGACCCAGCTCATTGCCGAC 2120
Db 217690 GAGGGCTTCGACAGGCGCCGATCTGTGCTGCCACTGGGGTTACCGGCTCGGTTTGCC 217749

Qy  2121 GTGGCCGCCGCCGGAACCCAAACACCGTCGCGTGCTGCTCAGTGCAGAGACGAGAGGAGATG 2180
Db 217750 GTCGCGTCCCGCCAAGCGAATACCGTTGTGGTCGTTGCTTGAGACGGGCCAACCATG 217809

Qy  2181 CCCTGGCTTCGACGGCACGCCCGGCCCGTCATCCAGGCCGTGGTGTACGCGGCAAACGAGACGGGC 2240
Db 217810 CCCTGGCCGTGCGGAGACTCGGTGAAGGAGCCATCATCCAGGCCGGGTTGGGGGC 217869

Qy  2241 AACTCCATTGCCGAGGTCGTGTCTTTGGGACTACACAACCCCTCGGGGCCAAGCGTGTCCCTCAGC 2300
Db 217870 CAGGCCGGTTGCGGAGAGTTGTGACCGGCAGTGAATTCCTTGCGGCCGGCGATCACC 217929

Qy  2301 TTCCCCAAGCGCCTGCAGGACAACCCCGGGCGGTTCTCAACTTCCGGCACCGAGC-GGCGGGCG 2359
Db 217930 TTCCCGCGGTCGGATCTCGGTCAGGACGGCCACCCGCTCGGAAACCGGCACGGGCTGGCCGGTGG 217989

Qy  2360 ------------------CACGCGGGTCTGGGCGGGGCGGGG 2405
Db 217990 GGGGACAATCGACCACCATCCACTCACCGAGGGCGGCGGC 218049

Qy  2406 TTTGCCGACAAGGAGGTCAATTTCCCCTTTGGCCGCACGGCCTGTCCTCCTACACACACTTTTTGCC 2465
Db 218050 AGCACAAATCAGACACCCCGGATGTTCGCGTTCGGGTTCGCACGGCGTTGTCCTATACCACAGTTTGCAG 218109

Qy  2466 TTTTCCAAATCTCTCCGTGTCTCACAAGGACACGGACAAGCGTGAGCGGTGTCCCTCTCCCGTCGAAG 2525
Db 218110 TATCGTGACCTGGGTGGTGACGGGC---GGCCACACCGTGCACGGCCACGTTTTCAGCGGTTACC 218166

Qy  2526 AACACCGGCGTCCGTGCCCGCGGCGCACAGGTGGCCCAGCTCTACGTCAAGCCCCTCCCAAGCG 2585
Db 218167 AACACGGGCGCCGACCGCGGGGCGGGATCGTCCCCAGCTGTATATGATCGGCAGCTCCGGC 218226

Qy  2586 GCCAAGAGATTAACCGCCCCGTCAGGGAGCTTCGCGAAAGGTCGAACTGCAGCGCC 2645
Db 218227 GAATCGCGGTTGCG------------GTTGCTGGGATTGAGCGGGATCGAGCGCGGACACCGGCGGTAC 218274

Qy  2646 GGCGAGGACGAAGGCGGTGACAATCGAGGACGACGAGAAGTACGTACGTCCGTCGT 2705
Db 218275 GGCCAGGACTCGGCGCCGGTAAGGATCGAGGCGGGACAACCCGGCGACTCCGGCGACCGG 218334

Qy  2706 GAGG 2709
         ||||
Db 218335 GAGG 218338
```

us-10-027-000-1.rng